AFOSR-TR-97 0700

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 1996	FINAL 01 SEP 94 to 31 AUG 95

4. TITLE AND SUBTITLE

1994 Conference on Neural Information Processing Systems - Natural and Synthetic

5. FUNDING NUMBERS

F49620-94-1-0394

6. AUTHOR(S)

Prof. John E. Moody

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Neural Information Processing System Foundation
P.O. Box 91000
Portland OR 97391-1000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NM
110 Duncan Avenue Suite B115
Bolling AFB DC 20332-8050

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release. Distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

In 1994 the Neural Information Processing System (NIPS) Foundation was awarded $15,000 to support travel awards for young researchers to attend the Neural Information Processing System Meeting held in Denver, Colorado, on November 28 to December 1, 1994. Since receiving this grant, the Neural Information Processing System Foundation Office has now moved to:
The Salk Institute for Biological Studies - CNL, 10010 North Torrey Pines Road, La Jolla CA 92037.

19971217 007

14. SUBJECT TERMS

Neural Information Processing

15. NUMBER OF PAGES
118

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

DTIC QUALITY INSPECTED 3

Neural Information Processing Systems

Student Travel Awards 1994

	Amount Awarded
Tony Bell Computational Neurobiology Laboratory The Salk Institute 10010 N. Torrey Pines Road La Jolla CA 92037-1009 USA	$600
Justin Boyan Carnegie Mellon University Comp. Science Dept. 5000 Forbes Ave. Pittsburgh PA 15213-3891 USA	600
Christoph Bregler UC Berkeley, ICSI 1947 Center St. #600 Berkeley CA 94704 USA	600
Richard Caruana Carnegie Mellon University School of Computer Science 500 Forbes Ave. Pittsburgh PA 15207 USA	600
David Cohn Massachusetts Institute of Technolgy Dept of Brain & Cognitive Sciences E10-243 Cambridge MA 02139 USA	600

	Amount Awarded
Trevor Darrell MIT Media Lab 20 Ames St. Room E15-388 Cambridge MA 02139 USA	600
Dr. Sreerupa Das University of Colorado Department of Computer Science Boulder CO 80309-0430 USA	250
Bard Ermentrout University of Pittsburgh Dept. of Mathematics Pittsburgh PA 15220 USA	600
Dr. Steven A. Gold Yale University Department of Computer Science P.O. Box 2158 Yale Station New Haven CT 06520-2158 USA	600
Hayit Greenspan California Institute of Technology M/S 116-81 Pasadena CA 91125 USA	250
Jchn Houde Massachusetts Institute of Technology 79 Amherst St. E10-105B Cambridge MA 02139 USA	250

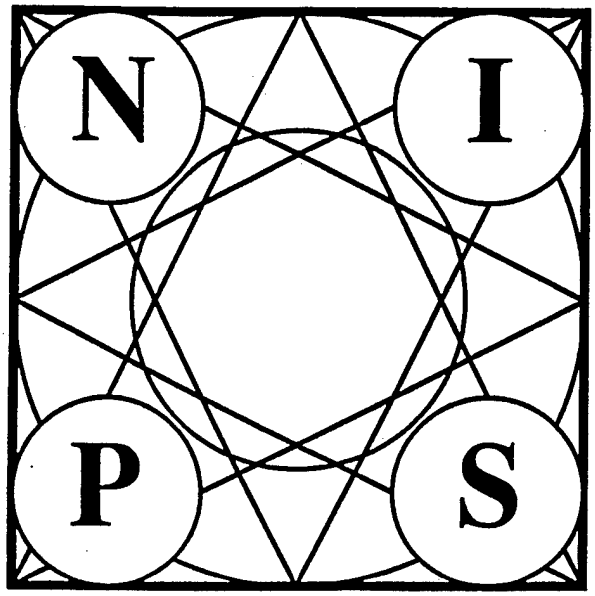
	Amount Awarded
Tommi Jaakkola Massachusetts Institute of Technology Dept Brain & Cognitive Sciences 79 Amherst St E10-105A Cambridge MA 02139 USA	600
Javier R. Movellan Dept of Cognitive Science University of California San Diego 9500 Gilman Drive DEPT 0515 La Jolla CA 92093-0515 USA	600
Genevieve Orr P O Box 91000 Portland OR 97291 USA	250
Klaus Pawelzik Salk Institute CNL P O Box 85800 San Diego CA 92138 USA	250
Alexandre Pouget 10010 No. Torrey Pines Rd. La Jolla CA 92037 USA	600
Rajesh P.N. Rao Department of Computer Science University of Rochester Rochester NY 14627 USA	600

	Amount Awarded
David Redish Carnegie Mellon University 5000 Forbes Ave Pittsburgh PA 15213 USA	250
Philip Sabes Massachusetts Institute of Technology E10-109 79 Amherst St Cambridge MA 02139 USA	250
Lawrence Saul Massachusetts Institute of Technology E10-243 Cambridge MA 02139 USA	600
Nicol N. Schraudolph Computational Neurobiology Laboratory The Salk Institute for Biological Sciences 10010 North Torrey Pines Road La Jolla CA 92037-1099 USA	600
Satinder Singh Massachusetts Institute of Technology Dept. of Brain and Cognitive Sciences Room E25-213a Cambridge MA 02139 USA	600

	Amount Awarded
Joseph Sirosh Department of Computer Sciences Taylor Hall 2124 University of Texas at Austin Austin TX 78712-11888	400
Randall R. Spangler California Institute of Technology Computation & Neural Systems 116-81 Pasadena CA 91125 USA	250
Magnus Stensmo P.O. Box 85800 San Diego CA 92186-5800 USA	600
Venkataraman Sundareswaran Intelligent Systems Laboratory College of Engineering Boston University 44 Cummington Street Boston MA 02215 USA	600
Joshua Tenenbaum Massachusetts Institute of Technology Department o Brain & Cognitive Science 77 Massachusetts Avenue E25-147B Cambridge MA USA	600
Emanuel V. Todorov Massachusetts Institute of Technology E25-147A Cambridge MA 02139 USA	600
Michael J. Turmon Room 393 E&TC Cornell University Ithaca NY 14853 USA	600

	Amount Awarded
Richard Zemel 600 Carnegie-Mellon University Department of Psychology Schenley Park Pittsburgh, PA 15213 USA	
Total Amount Awarded	**$15,000**

Abstracts Of Papers



1994 Conference on

NEURAL INFORMATION PROCESSING SYSTEMS – NATURAL AND SYNTHETIC

November 28 - December 1, 1994

Denver Marriott Hotel - City Center
Denver, Colorado

Workshops at Radisson Vail Resort
December 1- 3, 1994

8th Annual Conference on Neural Information Processing Systems Natural and Synthetic

November 28 - December 1, 1994

DENVER MARRIOT CITY CENTER
DENVER, COLORADO
WITH A POST-MEETING WORKSHOP, DECEMBER 1- DECEMBER 3
RADISSON RESORT, VAIL, COLORADO

Sponsored by NEURAL INFORMATION PROCESSING SYSTEMS Foundation, Inc. Additional sponsorship provided by Colorado School of Mines and Interval Research Corporation. Sponsorship of student and young investigator travel awards provided by Advanced Research Projects Agency and Office of Naval Research.

The progam stresses interdisciplinary interactions. All papers were thoroughly refereed. No parallel sessions. Roughly 35% of all papers submitted were accepted for oral or poster presentations.

ORGANIZING COMMITTEE:

General Chair	Gerald Tesauro, IBM
Program Chair	David Touretzky, Carnegie-Mellon
Workshop Chair	Todd Leen, Oregon Graduate Institute
Publicity Chair	Bartlett Mel, Caltech
Publications Chair	Joshua Alspector, Bellcore
Treasurer	Rodney Goodman, Caltech
Government/Corporate Liaison	John Moody, Oregon Graduate Institute
Local Arrangements Chair	Lori Pratt, Colorado School of Mines
Tutorials Co-Chairs	Steve Hanson, Siemens
	Gerald Tesauro, IBM
Contracts	Steve Hanson, Siemens
	Scott Kirkpatrick, IBM

PUBLICITY COMMITTEE:

Publicity Chair	Bartlett Mel, Caltech
Overseas Liaison (Japan)	Mitsuo Kawato, ATR Research Laboratories
Overseas Liaison (Australia, Singapore, India)	Marwan Jabri, University of Sydney
Overseas Liaison (Europe)	Joachim Buhmann, University of Bonn
Overseas Liaison (United Kingdom)	Alan Murray, Edinburgh University
Overseas Liaison (South America)	Andreas Meier, Simon Bolivar University

PROGRAM COMMITTEE:

Program Chair	David Touretzky, Carnegie Mellon
Program Co-Chairs	Subutai Ahmad, Interval Research
	Chuck Anderson, Colorado State University
	Eric Baum, NEC Research Institute
	Tom Brown, Yale University
	Scott Fahlman, Carnegie Mellon
	Terrence Fine, Cornell University
	Michael Jordan, MIT
	John Lazzaro, UC Berkeley
	Yann LeCun, AT&T Bell Labs
	Richard Lippmann, MIT
	Paul Munro, University of Pittsburgh
	John Platt, Synaptics

NIPS FOUNDATION BOARD MEMBERS

President	Terry Sejnowski
Vice President of Development	John Moody
Treasurer	Rodney Goodman
Secretary	Scott Kirkpatrick
Members	Jack Cowan
	Richard Lippmann
	Stephen Hanson
IEEE Representative	Terrence Fine
Legal Advisor	Philip K. Sotel

REVIEWERS:

Joshua Alspector, Bellcore
James Anderson, Brown University
Chris Atkeson, Georgia Tech
Jonathan Bachrach, IRCAM
Pierre Baldi, Caltech
Shumeet Baluja, Carnegie Mellon University
Etienne Barnard, Oregon Graduate Institute
Andrew R. Barron, Yale
Peter Bartlett, Australian National University
Andy Barto, University of Massachusetts
Sue Becker, McMaster University
Yoshua Bengio,University of Montreal
Uli Bodenhausen, University of Karlsruhe
Brian Bonnlander, University of Colorado, Boulder
Leon Bottou, Neuristique
Herve Bourlard, L&H Speech Products
Justin Boyan, Carnegie Mellon University
Jane Bromley, AT&T Bell Laboratories
Tim X. Brown, Bellcore
Tom Brown, Yale
Ken Buckland, PMC-Sierra, Inc.
Joachim Buhmann, University of Bonn
Wray Buntine, NASA, Ames
David J. Burr, Bellcore
William Byrne, Entropic Research Laboratory
N.T. Carnevale, Yale
Tzi-dar Chiueh, National Taiwan University
David Chalmers, Washington University
Michael Cohen, SRI International
David Cohn, MIT
Gary Cottrell, University of California, San Diego
Chris Darken, Siemens
Shawn Day, Synaptics Inc.
Dave DeMers, UC San Diego
Thomas G. Dietterich, Oregon State University
Georg Dorffner, University of Vienna
Gerard Dreyfus, ESPCI Laboratoire d'Electronique
Harris Drucker, AT&T Bell Laboratories
Richard O. Duda, SRI International
Mark Fanty, Oregon Graduate Institute
Bernd Fritzke, Ruhr University, Bochum
Patrick Gallinari, University of Paris
Michael Gasser, Indiana University
Zoubin Ghahramani, MIT
C. Lee Giles, NEC Research Institute
Dave Gillespie, Synaptics Inc.
Gene Gindi, Yale
Frederico Girosi, MIT
Rodney Goodman, Caltech
Hans-Peter Graf, AT&T Bell Laboratories
Patrick Haffner, Centre National d"Etudes des Telecommunications
Dan Hammerstrom, Adaptive Solutions
John Hampshire, Jet Propulsion Laboratory
Catherine Harris, Boston University
John Harris, University of Florida
Sherif Hashem, Pacific Northwest Laboratories
Michael Hasselmo, Harvard University
Babak Hassibi, Stanford
David Haussler, University of California,Santa Cruz
Simon Haykin, McMaster University
Andreas Herz, Beckman Institute
Geoffrey Hinton, University of Toronto
Bill Horne, NEC
John Houde, MIT
Nathan Intrator, Brown University
Marwan Jabri, Sydney University
Robert Jacobs, University of Rochester
Chuanyi Ji, Rensselaer Polytechnic Institute
Dan Johnston, Baylor College of Medicine
B.H. Juang, AT&T Bell Laboratories
Stephen Judd, Siemens Corporate Research
Visakan Kadirkamanathan, University of Sheffield
Michael Kearns, AT&T Bell Laboratories
Dan Kersten, Max-Planck Institute
Christof Koch, Caltech
Phil Kohn, ICSI
Anders Krogh, Technical University of Denmark
Anthony Kuh, University of Hawaii
Gary Kuhn, Siemens Corporate Research
S.Y. Kung, Waseda University
Kevin Lang, NEC
Harold Levy, Caltech
Long-Ji Lin, Siemens
Tsungnan Lin, Princeton
James Little, University of British Columbia
Michael Littman, Brown University
Marco Maggini, Universita di Firenze
Eve Marder, Brandeis University
Scott A. Markel, David Samoff Research Center
George Mamellos, Yale
Bimal Mathur, Rockwell International
Bartlett Mel, Caltech
Risto Miikkulainen, The University of Texas,Austin
Kenneth Miller, University of California, San Francisco
Melanie Mitchell, Sante Fe Institute
Eric Mjolsness, Yale University
Martin Moller, Aarhus University
Andrew Moore, Carnegie Mellon University
Nelson Morgan, ICSI
Anthony Movshon, New York University
Bob Narendra, Yale

Steven Nowlan, Synaptics
Bruno Olshausen, Cornell
Steve M Omohundro, ICSI
Alice O'Toole, University of Texas at Dallas
Satinder Pal Singh, MIT
Kannan Parthasarathy, Lexicus
Barak Pearlmutter, Siemens Coporate Research
Carsten Peterson, University of Lund
Jim Peterson, Clemson
Tom Petsche, Siemens
Tony Plate, University of British Columbia
John Platt, Synaptics
David Plaut, Carnegie Mellon University
Mark Plutowski, University of California San Diego
Jordan Pollack, Ohio State University
Dean Pomerleau, Carnegie Mellon University.
Lorien Pratt, Colorado School of Mines
Jose C. Principe, University of Florida
Mazin Rahim, AT&T Bell Laboratories
Anand Rangarajan, Yale
Steve Redman, Australian National University
A. David Redish, Carnegie Mellon University
Steve Renals, Cambridge University
Tony Robinson, University of Cambridge
Vwani Roychowdhury, Purdue
Michael Rudnick, Tulane
Virginia de Sa, University of Rochester
Philip Sabes, MIT
Eduard Sackinger, AT&T Bell Laboratories
Terry Sanger, Jet Propulsion Laboratory
Warren S. Sarle, SAS Institute
Larry Saul, MIT
Eric Saund, Xerox PARC
Terrence Sejnowski, The Salk Institute
Dan Seligson, Intel Corp.
Jude Shavlik, University of 'Wisconsin
Hava Siegelmann, Technion
K-Y Siu, University of California, Irvine
Patrice Simard, AT&T Bell Laboratories
Massimo Sivilotti, Tanner Research
Steve Smith, Carnegie Mellon University
Roger Smith, Yale
Paul Smolensky, University of Colorado,Boulder
Padhraic Smyth, Jet Propulsion Laboratory
Sara Solla, Niels Bohr Institute
David Standley, Rockwell International
Mark St. John, University of California, San Diego
Paul Stolorz, Jet Propulsion Laboratory
David G. Stork, Ricoh
Rich Sutton, GTE Laboratories
Sebastian Thrun, University of Bonn
Rob Tibshirani, University of Toronto
Naftali Tishby, The Hebrew University
Volker Tresp, Siemens AG
Ah Chung Tsoi, University of Queensland
Michael Turmon, Cornell
Lyle Ungar, University of Pennsylvania
Vladimir Vapnik, AT&T Bell Laboratories
Bert de Vries, David Samoff Research Center
Kelvin Wagner, University of Colorado
Alex Waibel, Carnegie Mellon University
Chris Watkins, Springfield, Inc. (UK)
Steve Watkins, PCSI
Raymond Watrous, Siemens Corporate Research
John Wawrzynek, UC Berkeley
Andreas Weigend, University of Colorado
Janet Wiles, University ofQueensland
Chris Williams, University of Toronto
Ronald J. Williams, Northeastern University
Robert Williamson, Australian National University
Charles Wilson, University of Tennesee
David Wolpert, The Sante Fe Institute
Lei Xu, Chinese University of Hong Kong
Richard Zemel, The Salk Institute

PROGRAM HIGHLIGHTS

SUNDAY NOVEMBER 27

6:00 PM-10:00 PM REGISTRATION

MONDAY NOVEMBER 28

8:30 AM-6:00 PM REGISTRATION

9:30 AM-5:30 PM TUTORIALS

6:30 PM RECEPTION & CONFERENCE BANQUET

8:30 PM THREE BEDROOMS, TWO BATHS, AND FOUR HIDDEN LAYERS: THE NEURAL NET ADAPTIVE HOUSE (BANQUET TALK)
MICHAEL MOZER, University of Colorado, Boulder

TUESDAY NOVEMBER 30

8:30 AM ORAL SESSION 1
COGNITIVE NEUROSCIENCE

THE PROBLEM OF VISUAL AWARENESS (INVITED TALK)
F.H. CRICK, Salk Institute

9:20 AM POSTER PREVIEW SESSION I

10:40 AM ORAL SESSION 2
REINFORCEMENT LEARNING

11:20 AM POSTER PREVIEW SESSION II

2:00 PM ORAL SESSION 3
NEUROSCIENCE

SEEING AND DECIDING: A WINNER-TAKE ALL DECISION PROCESS IN THE CEREBRAL CORTEX (INVITED TALK)
W.T. NEWSOME, Stanford University School of Medicine

2:50 PM POSTER PREVIEW SESSION III

4:15 PM ORAL SESSION 4
LEARNING THEORY

4:55 PM POSTER PREVIEW SESSION IV

7:30 PM REFRESHMENTS AND POSTER SESSION 1

WEDNESDAY NOVEMBER 30

8:30 AM ORAL SESSION 5
APPLICATIONS

HANDWRITING RECOGNITION FOR THE NEWTON (INVITED TALK).
LEONID KITAINIK, ParaGraph International

9:20 AM POSTER PREVIEW SESSION V

10:40 AM ORAL SESSION 6
IMPLEMENTATION

11:20 AM POSTER PREVIEW SESSION VI

2:00 PM ORAL SESSION 7
SPEECH AND SIGNAL PROCESSING

2:00 PM CORRELOGRAMS: A TOOL FOR SOUND SEPARATION (INVITED TALK).
MALCOLM SLANEY, Apple Computer

2:50 PM POSTER PREVIEW SESSION VII

4:15 PM ORAL SESSION 8
VISION

4:35 PM POSTER PREVIEW SESSION VIII

7:30 PM REFRESHMENTS AND POSTER SESSION II

THURSDAY DECEMBER 1

8:30 AM ORAL SESSION 9
ALGORITHMS AND ARCHITECTURES

8:30 AM FINANCIAL APPLICATIONS OF LEARNING FROM HINTS (INVITED TALK)
YASIR S. ABU-MOSTAFA, Caltech

10:30 AM ORAL SESSION 10
ALGORITHMS AND ARCHITECTURES

11:50 PM ADJOURN TO VAIL FOR WORKSHOPS

5:00 PM REGISTRATION, RECEPTION AT VAIL

FRIDAY DECEMBER 2

7:30-9:30 WORKSHOP MORNING SESSION

4:30-6:30 WORKSHOP AFTERNOON SESSION

7:00 PM WORKSHOP BANQUET

SATURDAY DECEMBER 3

7:30-9:30 WORKSHOP MORNING SESSION

4:30-6:30 WORKSHOP AFTERNOON SESSION

7:00 PM WORKSHOP WRAP-UP MEETING

WORKSHOP SCHEDULE

FRIDAY DECEMBER 2

Schwaber et. al.	Novel Control Techniques from Biological Inspiration
Baldi et. al.	Machine Learning Approaches in Computational Molecular Biology
Petsche et. al.	Novelty Detection and Adaptive System Monitoring
Hermansky/Pavel	Anthropomorphic Speech Signal Processing
Sirosh	Computational Role of Lateral Connections in the Cortex
Lei Xu et. al.	Unsupervised Learning Rules and Visual Processing
Cottrell	Statistical and Neural Network Approaches to Natural Languauge Processing
Keller	Neural Networks in Medicine
Siegelmann	Advances in Recurrent Networks

SATURDAY DECEMBER 3

Mitchell	Open and Closed Problems in Neural Network Robotics
Back/Wan	Neural Network Architectures with Time Delay Connections for Nonlinear Signal Processing
Perrone	Algorithms for High Dimensional Space: What Works and Why
Bishop	Doing it Backwards: Neural Networks and the Solution of Inverse Problems
Ermentrout	The Neural Basis of Locomotion: Models of Pattern Generators
Lei Xu et. al.	Unsupervised Learning Rules and Visual Processing
Cottrell	Statistical and Neural Network Approaches to Natural Languauge Processing
Keller	Neural Networks in Medicine
Siegelmann	Advances in Recurrent Networks

TUTORIAL PROGRAM

November 28, 1994

Session I: 9:30-11:30 am

9:30-11:30 *Recent Advances in Learning Theory*
Michael Kearns, AT&T Bell Laboratories 25

9:30-11:30 *A Survey of Pattern Recognition Hardware*
Dan Hammerstrom, OGI and Adaptive Solutions Inc. 26

Session II: 1:00-3:00 pm

1:00-3:00 *Advances in the Theory and Applications of the Self-Organizing Map*
Teuvo Kohonen, Helsinki University of Technology 26

1:00-3:00 *Learning to Act: An Introduction to Reinforcement Learning*
Andy Barto, University of Massachusetts at Amherst 27

Session III: 3:30-5:30 pm

3:30-5:30 *Images of the Mind: A Tutorial on Brain Imaging*
Marc Raichle, Washington University Medical School (St. Louis) 28

3:30-5:30 *Statistics and Nets: Understanding Nonlinear Models from Their Linear Relatives*
Leo Breiman, University of California at Berkeley 28

TUESDAY AM

ORAL SESSION 1
COGNITIVE NEUROSCIENCE

8:30 *O1.1* *THE PROBLEM OF VISUAL AWARENESS (INVITED TALK)*
F.H. CRICK, Salk Institute. 29

9:00 *O1.2* *DIRECTION SELECTIVITY IN PRIMARY VISUAL CORTEX USING MASSIVE INTRACORTICAL CONNECTIONS*
HUMBERT SUAREZ, Caltech, CHRISTOF KOCH, Caltech, and RODNEY DOUGLAS, University of Oxford. 29

9:20 *SPOTLIGHT I: COGNITIVE NEUROSCIENCE*
PLASTICITY AS LIKELIHOOD OF RELEVANCE: COMPETITION IN DISTRIBUTED REPRESENTATIONS, Nicol N. Schraudolph and Terrence J. Sejnowski, Computational Neurobiology Laboratory, Salk Institute 30

GRAMMAR LEARNING BY A SELF-ORGANIZING NETWORK, Michiro Negishi, Boston University 30

PATTERNS OF DAMAGE IN NEURAL NETWORKS: THE EFFECTS OF LESION AREA, SHAPE AND NUMBER,Eytan Ruppin and James A Reggia, University of Maryland 30

9:30 *O1.3* *ON THE COMPUTATIONAL UTILITY OF CONSCIOUSNESS*
DONALD W. MATHIS AND MICHAEL C. MOZER, University of Colorado 30

9:50 *O1.4* *TEMPORAL CHARACTERISTICS OF DYNAMIC MOTOR LEARNING*
TOM BRASHERS-KRUG, EMANUEL V. TODOROV, and REZA SHADMEHR, MIT 30

10:10 BREAK

ORAL SESSION 2
REINFORCEMENT LEARNING

10:40 O2.1 REINFORCEMENT LEARNING ALGORITHM FOR PARTIALLY OBSERVABLE MARKOV DECISION PROBLEMS
TOMMI JAAKKOLA, SATINDER P. SINGH, and MICHAEL I. JORDAN, MIT 31

11:00 O2.2 ADVANTAGE UPDATING APPLIED TO A DIFFERENTIAL GAME
MANCE E. HARMON, LEEMON C. BAIRD III, and A. HARRY KLOPF, Wright Laboratory 31

11:20 SPOTLIGHT II: REINFORCEMENT LEARNING
OPTIMAL MOVEMENT PRIMITIVES,Terence Sanger, Jet Propulsion Laboratory 32

AN INTEGRATED ARCHITECTURE OF ADAPTIVE NEURAL NETWORK CONTROL FOR DYNAMIC SYSTEMS, Liu Ke, Robert L. Tokar,and Brian D. McVey, Los Alamos National Laboratory 32

PHASE-SPACE LEARNING, Fu-Sheng Tsung and Garrison W. Cottrell, University of California, San Diego 32

11:30 02.3 REINFORCEMENT LEARNING WITH SOFT STATE AGGREGATION
SATINDER P. SINGH, TOMMI JAAKKOLA, and MICHAEL I. JORDAN, MIT 32

11:50 02.4 GENERALIZATION IN REINFORCEMENT LEARNING: SAFELY APPROXIMATING THE VALUE FUNCTION
JUSTIN A. BOYAN and ANDREW W. MOORE, Carnegie Mellon University 32

12:00 LUNCH

TUESDAY PM

ORAL SESSION 3
NEUROSCIENCE

2:00 O3.1 SEEING AND DECIDING: A WINNER-TAKE-ALL DECISION PROCESS IN THE CEREBRAL CORTEX(INVITED TALK)
W.T. NEWSOME, Stanford University School of Medicine 33

2:30 O3.2 A MODEL FOR CHEMOSENSORY RECEPTION
RAINER MALAKA and THOMAS RAGG, Universitat Karlsruhe, and MARTIN HAMMER, Freie Universitat Berlin 33

2:50 SPOTLIGHT III: NEUROSCIENCE
MODEL OF A BIOLOGICAL NEURON AS A TEMPORAL NEURAL NETWORK, Sean D. Murphy and Edward W. Kairiss, Yale University 33

A CRITICAL COMPARISON OF MODELS FOR ORIENTATION AND OCULAR DOMINANCE COLUMNS IN THE STRIATE CORTEX, Ed Erwin and Klaus Obermayer, Universitat Bielefeld 33

A NOVEL REINFORCEMENT MODEL OF BIRDSONG VOCALIZATION LEARNING, Kenji Doya and Terrence J. Sejnsowski, Howard Hughes Medical Institute, Salk Institute 33

3:00 *O3.3* *THE ELECTRONIC TRANSFORMATION: A TOOL FOR RELATING NEURONAL FORM TO FUNCTION*
NICHOLAS T. CARNEVALE, KENNETH Y. TSAI, AND THOMAS H. BROWN, Yale University and BRENDA J. CLAIBORNE, University of Texas
33

3:20 *FEEDBACK REGULATION OF CHOLINERGIC MODULATION AND AUTO-ASSOCIATIVE MEMORY FUNCTION IN HIPPOCAMPAL REGION CA3*
MICHAEL E. HASSELMO, EDI BARKAI and JOSHUA BERKE, Harvard University
34

3:40 *BREAK*

ORAL SESSION 4
LEARNING THEORY

4:15 *O4.1* *ON THE COMPUTATIONAL COMPLEXITY OF NETWORKS OF SPIKING NEURONS*
WOLFGANG MAASS, Technische Universitaet Graz, Austria
34

4:35 *O4.2* *OPTIMAL TRAINING ALGORITHMS AND THEIR RELATION TO BACKPROPAGATION*
BABAK HASSIBI and THOMAS KAILATH, Stanford University
34

4:55 *SPOTLIGHT IV: LEARNING THEORY*
RESPONSE FUNCTIONS FOR LEARNING IN LARGE LINEAR PERCEPTRONS, Peter Sollich, University of Edinburgh 35

GENERALISATION IN FEEDFORWARD NETWORKS, Adam Kowalczyk and Herman Ferra, Telecom Australia, Research Laboratories 35

FROM DATA DISTRIBUTIONS TO REGULARIZATION IN INVARIANT LEARNING, Todd K. Leen, Oregon Graduate Institute of Science and Technology
35

NEURAL NETWORK ENSEMBLES, CROSS VALIDATION, AND ACTIVE LEARNING, Anders Krogh and Jesper Vedelsby, Technical University of Denmark
35

5:10 *O4.3* *SYNCHRONY AND DESYNCHRONY IN OSCILLATOR NETWORKS*
DE LIANG WANG AND DAVID TERMAN,Ohio State University
35

5:30 *DINNER*

7:30 *REFRESHMENTS AND POSTER SESSION I*

TUESDAY EVENING POSTERS

ALGORITHMS & ARCHITECTURES

7:30 *AA:1* *EXTRACTING RULES FROM ARTIFICIAL NEURAL NETWORKS WITH DISTRIBUTED REPRESENTATIONS*
SEBASTIAN THRUN, University of Bonn 36

7:30 *AA:2* *CAPACITY AND INFORMATION EFFICIENCY OF A BRAIN-LIKE ASSOCIATIVE NET*
BRUCE GRAHAM and DAVID WILLSHAW 36

7:30 AA:3 *BOOSTING THE PERFORMANCE OF RBF NETWORKS WITH DYNAMIC DECAY ADJUSTMENT*
MICHAEL R. BERTHOLD, Forschungszentrum Informatik and JAY DIAMOND, Intel Corp. 37

7:30 AA:4 *SIMPLIFYING NETWORKS BY DISCOVERING "FLAT" MINIMA*
SEPP HOCHREITER and JURGEN SCHMIDHUBER, Technische Universitat Munchen 37

7:30 AA:5 *LEARNING WITH PRODUCT UNITS*
LAURENS R. LEERINK and MARWAN A. JABRI, University of Sydney, and C. LEE GILES and BILL G. HORNE, NEC Research Institute 37

7:30 AA:6 *DETERMINISTIC ANNEALING VARIANT OF THE EM ALGORITHM*
NAONORI UEDA and RYOHEI NAKANO, NTT Communication Science Laboratories 38

7:30 AA:7 *PLASTICITY AS LIKELIHOOD OF RELEVANCE: COMPETITION IN DISTRIBUTED REPRESENTATIONS*
NICOL N. SCHRAUDOLPH and TERRENCE J. SEJNOWSKI, Salk Institute 38

7:30 AA:8 *DIFFUSION OF CREDIT IN MARKOVIAN MODELS*
YOSHUA BENGIO, Universite de Montreal and PAOLO FRASCONI, Universita de Firenze, Italy 38

7:30 AA:9 *MIXTURE OF ONE-DIMENSIONAL PROJECTIONS (MODP): A UNIFYING ARCHITECTURE FOR PRINCIPAL COMPONENT ANALYSIS AND COMPETITIVE LEARNING*
JOSHUA B. TENENBAUM AND EMANUEL V. TODOROV, MIT 38

7:30 AA:10 *INTERIOR POINT IMPLEMENTATIONS OF ALTERNATING MINIMIZATION TRAINING*
MICHAEL LEMMON and PETER T. SZYMANSKI, University of Notre Dame 39

7:30 AA:11 *SARDNET: A SELF-ORGANIZING FEATURE MAP FOR SEQUENCES*
DANIEL L. JAMES and RISTO MIIKKULAINEN, University of Texas at Austin 39

7:30 AA:12 *CONVERGENCE PROPERTIES OF THE K-MEANS ALGORITHMS*
LEON BOTTOU, Neuristique and YOSHUA BENGIO, Universite de Montreal 39

7:30 AA:13 *ACTIVE LEARNING FOR FUNCTION APPROXIMATION*
KAH KAY SUNG and PARTHA NIYOGI, MIT 39

7:30 AA:14 *PHASE-SPACE LEARNING*
FU-SHENG TSUNG and GARRISON W. COTTRELL, University of California, San Diego 39

7:30 AA:15 *ANALYSIS OF UNSTANDARDIZED CONTRIBUTIONS IN CROSS CONNECTED NETWORKS*
THOMAS R. SHULTZ, YURIKO OSHIMA-TAKANE, and YOSHIO TAKANE, McGill University 40

7:30 AA:16 *TEMPLATE-BASED ALGORITHMS FOR CONNECTIONIST RULE EXTRACTION*
JAY A. ALEXANDER and MICHAEL C. MOZER, University of Colorado 40

COGNITIVE SCIENCE

7:30 CS:1 *GRAMMAR LEARNING BY A SELF-ORGANIZING NETWORK*
MICHIRO NEGISHI, Boston University 40

7:30 CS:2 *FORWARD DYNAMIC MODELS IN HUMAN MOTOR CONTROL: PSYCHOPHYSICAL EVIDENCE*
DANIEL M. WOLPERT, ZOUBIN GHAHRAMANI and MICHAEL I. JORDAN, MIT 41

7:30 CS:3 *PATTERNS OF DAMAGE IN NEURAL NETWORKS: THE EFFECTS OF LESION AREA, SHAPE AND NUMBER*
EYTAN RUPPIN and JAMES A. REGGIA, University of Maryland 41

CONTROL

7:30 CN:1 *OPTIMAL MOVEMENT PRIMITIVES*
TERENCE D. SANGER, Jet Propulsion Laboratory 41

7:30 CN:2 *AN INTEGRATED ARCHITECTURE OF ADAPTIVE NEURAL NETWORK CONTROL FOR DYNAMIC SYSTEMS*
LIU KE, ROBERT L. TOKAR, and BRIAN D. McVEY, Los Alamos National Laboratory 42

IMPLEMENTATIONS

7:30 IM:1 *PULSESTREAM SYNAPSES WITH NON-VOLATILE ANALOGUE AMORPHOUS-SILICON MEMORIES*
A.J. HOLMES, A.F. MURRAY, S. CHURCHER and J. HAJTO, University of Edinburgh, and M.J. ROSE, Dundee University 42

7:30 IM:2 *A LAGRANGIAN FORMULATION FOR TRAINING OF KERR-TYPE OPTICAL NETWORKS*
JAMES E. STECK, STEVEN R. SKINNER, and ELIZABETH C. BEHRMAN, The Wichita State University 42

7:30 IM:3 *A CHARGE-BASED CMOS PARALLEL ANALOG VECTOR QUANTIZER*
GERT CAUWENBERGHS, John Hopkins University and VOLNEI PEDRONI, California Institute of Technology 43

7:30 IM:4 *AN AUDITORY LOCALIZATION AND COORDINATE TRANSFORM CHIP*
TIMOTHY HORIUCHI, California Institute of Technology 43

LEARNING THEORY

7:30 LT:1 *HIGHER ORDER STATISTICAL DECORRELATION WITHOUT INFORMATION LOSS*
GUSTAVO DECO, Siemens, AG, and WILFRIED BRAUER, Technische Universitat Munchen 44

7:30 LT:2 *HYPERPARAMETERS, EVIDENCE AND GENERALISATION IN AN UNREALISABLE SCENARIO*
GLENN MARION and DAVID SAAD, University of Edinburgh 44

7:30 LT:3 *RESPONSE FUNCTIONS FOR LEARNING IN LARGE LINEAR PERCEPTRONS*
PETER SOLLICH, University of Edinburgh 44

7:30 LT:4 *GENERALIZATION DYNAMICS IN NEURAL NETWORKS*
CHANGFENG WANG and SANTOSH S. VENKATESH, University of Pennsylvania 45

7:30 LT:5 *STOCHASTIC DYNAMICS OF THREE-STATE NEURAL NETWORKS*
TORU OHIRA, Sony Computer Science Laboratory and JACK D. COWAN, University of Chicago
45

7:30 LT:6 *LEARNING STOCHASTIC PERCEPTRONS UNDER K-BLOCKING DISTRIBUTIONS*
MARIO MARCHAND and SAEED HADJIFARADJI, University of Ottawa
46

7:30 LT:7 *GENERALISATION IN FEEDFORWARD NETWORKS*
ADAM KOWALCZYK, Telecom Australia Research Laboratories.
46

7:30 LT:8 *FROM DATA DISTRIBUTIONS TO REGULARIZATION IN INVARIANT LEARNING*
TODD K. LEEN, Oregon Graduate Institute of Science and Technology
46

7:30 LT:9 *NEURAL NETWORK ENSEMBLES, CROSS VALIDATION, AND ACTIVE LEARNING*
ANDERS KROGH and JESPER VEDELSBY, Technical University of Denmark
47

NEUROSCIENCE

7:30 NS:1 *OCULAR DOMINANCE AND ACTIVATION DYNAMICS IN A UNIFIED SELF-ORGANIZING MODEL OF THE VISUAL CORTEX*
JOSEPH SIROSH and RISTO MIIKKULAINEN, University of Texas at Austin
47

7:30 NS:2 *ANATOMICAL ORIGIN AND COMPUTATIONAL ROLE OF DIVERSITY IN THE RESPONSE PROPERTIES OF CORTICAL NEURONS*
KALANIT GRILL SPECTOR, SHIMON EDELMAN and RAPHAEL MALACH, The Weizmann Institute of Science
47

7:30 NS:3 *MODEL OF A BIOLOGICAL NEURON AS A TEMPORAL NEURAL NETWORK*
SEAN D. MURPHY and EDWARD W. KAIRISS, Yale University
48

7:30 NS:4 *A CRITICAL COMPARISON OF MODELS FOR ORIENTATION AND OCULAR DOMINANCE COLUMNS IN THE STRIATE CORTEX*
ED ERWIN and KLAUS SCHULTEN, University of Illinois and KLAUS OBERMAYER, Universitat Bielefeld
48

7:30 NS:5 *A NOVEL REINFORCEMENT MODEL OF BIRDSONG VOCALIZATION LEARNING*
KENJI DOYA and TERRENCE J. SEJNOWSKI, Salk Institute
48

7:30 NS:6 *REINFORCEMENT LEARNING PREDICTS THE SITE OF PLASTICITY FOR AUDITORY REMAPPING IN THE BARN OWL*
ALEXANDRE POUGET, CEDRIC DEFFAYET, and TERRENCE J. SEJNOWSKI, Salk Institute
49

7:30 NS:7 *MORPHOGENESIS OF THE LATERAL GENICULATE NUCLEUS: HOW SINGULARITIES AFFECT GLOBAL STRUCTURE*
SVILEN TZONEV and KLAUS SCHULTEN, Beckman Institute, University of Illinois and JOSEPH G. MALPELI, University of Illinois
49

REINFORCEMENT LEARNING

7:30 RL:1 *INSTANCE-BASED STATE IDENTIFICATION FOR REINFORCEMENT LEARNING*
R. ANDREW McCALLUM, University of Rochester 50

7:30 RL:2 *FINDING STRUCTURE IN REINFORCEMENT LEARNING*
SEBASTIAN THRUN, University of Bonn and ANTON SCHWARTZ, Stanford University 50

7:30 RL:3 *REINFORCEMENT LEARNING METHODS FOR CONTINUOUS-TIME MARKOV DECISION PROBLEMS*
STEVEN J. BRADTKE and MICHAEL O. DUFF, University of Massachusetts 50

7:30 RL:4 *A CLASS OF ACTOR/CRITIC ARCHITECTURES THAT ARE EQUIVALENT TO Q-LEARNING*
ROBERT H. CRITES and ANDREW G. BARTO, University of Massachusetts 51

SPEECH AND SIGNAL PROCESSING

7:30 SP:1 *CONNECTIONIST SPEAKER NORMALIZATION WITH GENERALIZED RESOURCE ALLOCATING NETWORKS*
CESARE FURLANELLO, DIEGO GIULIANI, and EDMONDO TRENTIN, Istituto per La Ricerca Scientifica e Tecnologica 51

7:30 SP:2 *USING VOICE TRANSFORMATIONS TO CREATE ADDITIONAL TRAINING SPEAKERS FOR WORD SPOTTING*
ERIC I. CHANG and RICHARD P. LIPPMANN, MIT Lincoln Laboratory 51

7:30 SP:3 *A COMPARISON OF DISCRETE-TIME OPERATOR MODELS FOR NONLINEAR SYSTEM IDENTIFICATION*
ANDREW D. BACK and AH CHUNG TSOI, University of Queensland 52

VISION

7:30 VI:1 *JPMAX: LEARNING TO RECOGNIZE MOVING OBJECTS AS A MODEL-FITTING PROBLEM*
SUZANNA BECKER, McMaster University 52

7:30 VI:2 *PCA-PYRAMIDS FOR IMAGE COMPRESSION*
HORST BISCHOF and KURT HORNIK, Technical University Vienna 52

7:30 VI:3 *UNSUPERVISED CLASSIFICATION OF 3D OBJECTS FROM 2D VIEWS*
SATOSHI SUZUKI and HIROSHI ANDO, ATR Human Information Processing Research Laboratories 53

7:30 VI:4 *FAST ALGORITHMS FOR 2D AND 3D POINT MATCHING: POSE ESTIMATION AND CORRESPONDENCE*
STEVEN GOLD, CHIEN PING LU, ANAND RANGARAJAN, SUGUNA PAPPU, and ERIC MJOLSNESS, Yale University 53

WEDNESDAY AM

ORAL SESSION 5
APPLICATIONS

8:30 *O5.1* *HANDWRITING RECOGNITION FOR THE NEWTON (INVITED TALK)*
L. KITAINIK, ParaGraph International 54

9:00 *O5.2* *TRANSFORMATION INVARIANT AUTOASSOCIATION WITH APPLICATION TO HANDWRITTEN CHARACTER RECOGNITION*
HOLGER SCHWENK and MAURICE MILGRAM, Universite Pierre et Marie Curie 54

9:20 *SPOTLIGHT V: APPLICATIONS*
RECOGNIZING HANDWRITTEN DIGITS USING MIXTURES OF LINEAR MODELS, Geoffrey E. Hinton, Michael Revow, and Peter Dayan, University of Toronto 54

9:30 *O5.3* *LEARNING PROTOTYPE MODELS FOR TANGENT DISTANCE*
TREVOR HASTIE, PATRICE SIMARD, and EDUARD SACKINGER, AT&T Bell Laboratories 54

9:50 *O5.4* *REAL-TIME CONTROL OF A TOKAMAK PLASMA USING NEURAL NETWORKS*
CHRIS M. BISHOP, Aston University and PAUL S. HAYNES, MIKE E.U. SMITH, TOM N. TODD and DAVID L. TROTMAN, AEA Technology 55

10:10 *BREAK*

ORAL SESSION 6

IMPLEMENTATION

10:40 *O6.1* *ICEG MORPHOLOGY CLASSIFICATION USING AN ANALOGUE VLSI NEURAL NETWORK*
RICHARD COGGINS, MARWAN JABRI, BARRY FLOWER, and STEPHEN PICKARD, University of Sydney 55

11:00 *O6.2* *A SILICON AXON*
BRADLEY A. MINCH, PAUL HASLER, CHRIS DIORIO,and CARVER MEAD, California Institute of Technology 55

11:20 *SPOTLIGHT VI: IMPLEMENTATIONS*
PREDICTING THE RISK OF COMPLICATIONS IN CORONARY ARTERY BYPASS OPERATIONS USING NEURAL NETWORKS, Richard P. Lippmann and Yuchun Lee, MIT Lincoln Laboratory and Dr. David Shahian, Lahey Clinic 56

LOCAL ERROR BARS FOR NONLINEAR REGRESSION AND TIME SERIES PREDICTION, David A. Nix and Andreas S. Weigend, University of Colorado 56

DYNAMIC CELL STRUCTURES, Jorg Bruske and Gerald Sommer, Christian Albrechts University at Kiel, Germany 56

11:30 *O6.3* *THE NI1000: HIGH SPEED PARALLEL VLSI FOR IMPLEMENTING MULTILAYER PERCEPTRONS*
MICHAEL P. PERRONE, Thomas J. Watson Research Center and LEON N. COOPER, Brown University 56

11:50 *O6.4* *ANALOG VLSI IMPLEMENTATION OF THE ART1 ALGORITHM*
T. SERRANO, B. LINARES-BARRANCO, and J.L. HUERTAS, National Microelectronics Center, Spain 56

12:10 *LUNCH*

WEDNESDAY PM

ORAL SESSION 7
SPEECH AND SIGNAL PROCESSING

2:00 *O7.1* *CORRELOGRAMS : A TOOL FOR SOUND SEPARATION (INVITED TALK)*
M. SLANEY, Apple Computer 57

2:30 *O7.2* *NON-LINEAR PREDICTION OF ACOUSTIC VECTORS USING HIERARCHICAL MIXTURES OF EXPERTS*
S.R. WATERHOUSE and A.J. ROBINSON, Cambridge University 57

2:50 *SPOTLIGHT VII: SPEECH AND SIGNAL PROCESSING*

A CONNECTIONIST TECHNIQUE FOR ACCELERATED TEXTUAL INPUT: LETTING A NETWORK DO THE TYPING, Dean A. Pomerleau, Carnegie Mellon University 58

PREDICTIVE CODING WITH NEURAL NETS: APPLICATION TO TEXT COMPRESSION, Stefan Heil and Jurgen Schmidhuber, Technische Universitat Munchen 58

HIERARCHICAL MIXTURES OF EXPERTS APPLIED TO A FRAME-BASED NEURAL NETWORK SYSTEM FOR CONTINUOUS SPEECH RECOGNITION, Ying Zhao, Richard Schwartz and John Makhoul, BBN System and Technologies 58

3:00 *O7.3* *GLOVE-TALK II: MAPPING HAND GESTURES TO SPEECH USING NEURAL NETWORKS*
S. SIDNEY FELS and GEOFFREY E. HINTON, University of Toronto 58

3:20 *O7.4* *VISUAL SPEECH RECOGNITION WITH STOCHASTIC NETWORKS*
JAVIER R. MOVELLAN, University of California San Diego. 58

3:40 *BREAK*

ORAL SESSION 8
VISION

4:15 *O8.1* *LEARNING SACCADIC EYE MOVEMENTS USING MULTISCALE SPATIAL FILTERS*
RAJESH P.N. RAO and DANA H. BALLARD, University of Rochester 59

4:35 *SPOTLIGHT VIII: VISION*

LEARNING DIRECTION IN GLOBAL MOTION: TWO CLASSES OF PSYCHOPHYSICALLY-MOTIVATED MODELS, V. Sundareswaran and Lucia M. Vaina, Boston University 59

DECORRELATION DYNAMICS: A THEORY FOR ORIENTATION CONTRAST AND ADAPTATION, Dawei W. Dong, University of California, Berkeley 59

LIMITS ON LEARNING MACHINE ACCURACY IMPOSED BY DATA QUALITY, Corinna Cortes, L.D. Jackel, and Wan-Ping Chiang, AT&T Bell Laboratories 59

4:45 *O8.2* *A CONVOLUTIONAL NEURAL NETWORK HAND TRACKER*
STEVEN J. NOWLAN and JOHN C. PLATT, Synaptics, Inc. 59

5:05 O8.4 *CORRELATION AND INTERPOLATION NETWORKS FOR REAL-TIME EXPRESSION ANALYSIS/SYNTHESIS*
TREVOR DARRELL, IRFAN ESSA, and ALEX PENTLAND, MIT Media Lab
60

5:25 *DINNER*

WEDNESDAY EVENING POSTERS

ALGORITHMS & ARCHITECTURES

7:30 AA:21 *FACTORIAL LEARNING AND THE EM ALGORITHM*
ZOUBIN GHAHRAMANI, MIT
61

7:30 AA:22 *A GROWING NEURAL GAS NETWORK LEARNS TOPOLOGIES*
BERND FRITZKE, Ruhr-Universitat Bochum
61

7:30 AA:23 *LOCAL ERROR BARS FOR NONLINEAR REGRESSION AND TIME SERIES PREDICTION*
DAVID A. NIX and ANDREAS S. WEIGEND, University of Colorado
61

7:30 AA:24 *AN ALTERNATIVE MODEL FOR MIXTURES OF EXPERTS*
LEI XU, The Chinese University of Hong Kong, MICHAEL I. JORDAN, MIT, and GEOFFREY E. HINTON, University of Toronto
61

7:30 AA:25 *ESTIMATING CONDITIONAL PROBABILITY DENSITIES FOR PERIODIC VARIABLES*
CHRIS M. BISHOP and CLAIRE LEGLEYE, Aston University
62

7:30 AA:26 *EFFECTS OF NOISE ON CONVERGENCE AND GENERALIZATION IN RECURRENT NETWORKS*
KAM JIM, BILL G. HORNE and C. LEE GILES, NEC Research Institute
62

7:30 AA:27 *LEARNING MANY RELATED TASKS AT THE SAME TIME WITH BACKPROPAGATION*
RICH CARUANA, Carnegie Mellon University
62

7:30 AA:28 *A RAPID GRAPH-BASED METHOD FOR ARBITRARY TRANSFORMATION INVARIANT PATTERN CLASSIFICATION*
ALESSANDRO SPERDUTI, Universita di Pisa and DAVID G. STORK, Ricoh California Research Center
62

7:30 AA:29 *RECURRENT NETWORKS: SECOND ORDER PROPERTIES AND PRUNING*
MORTEN WITH PEDERSEN and LARS KAI HANSEN, Technical University of Denmark
63

7:30 AA:30 *CLASSIFYING WITH GAUSSIAN MIXTURES, CLUSTERS, AND SUBSPACES*
NANDA KAMBHATLA and TODD K. LEEN, Oregon Graduate Institute of Science & Technology
63

7:30 AA:31 *EFFICIENT METHODS FOR DEALING WITH MISSING DATA IN SUPERVISED LEARNING*
VOLKER TRESP, RALPH NEUNEIER and SUBUTAI AHMAD, Siemens AG
63

7:30 AA:32 *AN EXPERIMENTAL COMPARISON OF RECURRENT NEURAL NETWORKS*
BILL G. HORNE and C. LEE GILES, NEC Research Institute
63

7:30 AA:33 ACTIVE LEARNING WITH STATISTICAL MODELS
DAVID A. COHN, ZOUBIN GHAHRAMANI, and MICHAEL I. JORDAN, MIT 63

7:30 AA:34 DYNAMIC CELL STRUCTURES
JORG BRUSKE and GERALD SOMMER, Christian Albrechts University at Kiel 64

7:30 AA:35 LEARNING WITH PREKNOWLEDGE: CLUSTERING WITH POINT AND GRAPH MATCHING DISTANCE MEASURES
STEVEN GOLD, ANAND RANGARAJAN and ERIC MJOLSNESS, Yale University 64

7:30 AA:36 SETTLING TEMPORAL DIFFERENCES: TIME SERIES PREDICTION USING TD
PETER T. KAZLAS and ANDREAS S. WEIGEND, University of Colorado 64

APPLICATIONS

7:30 AP:21 COMPARING THE PREDICTION ACCURACY OF ARTIFICIAL NEURAL NETWORKS AND OTHER STATISTICAL MODELS FOR BREAST CANCER SURVIVAL
HARRY B. BURKE, DAVID B. ROSEN, and PHILIP H. GOODMAN, University of Nevada School of Medicine 65

7:30 AP:22 A CONNECTIONIST TECHNIQUE FOR ACCELERATED TEXTUAL INPUT: LETTING A NETWORK DO THE TYPING
DEAN A. POMERLEAU, Carnegie Mellon University 65

7:30 AP:23 LEARNING TO PLAY THE GAME OF CHESS
SEBASTIAN THRUN, University of Bonn 66

7:30 AP:24 PREDICTIVE CODING WITH NEURAL NETS: APPLICATION TO TEXT COMPRESSION
STEFAN HEIL and JURGEN SCHMIDHUBER, Technische Universitat Munchen 66

7:30 AP:25 PREDICTING THE RISK OF COMPLICATIONS IN CORONARY ARTERY BYPASS OPERATIONS USING NEURAL NETWORKS
RICHARD P. LIPPMANN and YUCHUN LEE, MIT Lincoln Laboratory and DR. DAVID SHAHIAN, Lahey Clinic 66

7:30 AP:26 A MIXTURE MODEL NEURAL EXPERT SYSTEM FOR DIAGNOSIS
MAGNUS STENSMO and TERRENCE J. SEJNOWSKI, Salk Institute 66

7:30 AP:27 INFERRING GROUND TRUTH FROM SUBJECTIVE LABELLING OF VENUS RADAR IMAGES
P. SMYTH, M. BURL, U.M. FAYYAD, P. BALDI, Jet PropulsionLaboratory and P. PERONA, California Institute of Technology 67

CHARACTER RECOGNITION

7:30 CR:21 THE USE OF DYNAMIC WRITING INFORMATION IN A CONNECTIONIST ON-LINE CURSIVE HANDWRITING RECOGNITION SYSTEM
STEFAN MANKE and MICHAEL FINKE, University of Karlsruhe, and ALEX WAIBEL, Carnegie Mellon University 67

7:30 CR:22 ADAPTIVE ELASTIC INPUT FIELD FOR RECOGNITION IMPROVEMENT
MINORU ASOGAWA, C&C Systems Research Laboratories, NEC 68

7:30 CR:23 *RECOGNIZING HANDWRITTEN DIGITS USING MIXTURES OF LINEAR MODELS*
GEOFFREY E. HINTON, MICHAEL REVOW, and PETER DAYAN, University of Toronto 68

7:30 CR:24 *PAIRWISE NEURAL NETWORK CLASSIFIERS WITH PROBABILISTIC OUTPUTS*
DAVID PRICE, STEFAN KNERR,LEON PERSONNAZ, and GERARD DREYFUS, ESPCI 68

CONTROL

7:30 CN:21 *FORMATION OF INTERNAL MODELS FOR LEARNING CONTROL OF ARM MOVEMENTS*
REZA SHADMEHR, TOM BRASHERS-KRUG, and FERDINANDO MUSSA-IVALDI, MIT. 69

7:30 CN:22 *COMPUTATIONAL STRUCTURE OF COORDINATE TRANSFORMATIONS: A GENERALIZATION STUDY*
ZOUBIN GHAHRAMANI, DANIEL M. WOLPERT, and MICHAEL I. JORDAN, MIT 69

COGNITIVE SCIENCE

7:30 CS:21 *A SOLVABLE CONNECTIONIST MODEL OF IMMEDIATE RECALL OF ORDERED LISTS*
NEIL BURGESS, UCL, London 70

IMPLEMENTATIONS

7:30 IM:21 *AN ANALOG NEURAL NETWORK INSPIRED BY FRACTAL BLOCK CODING*
FERNANDO J. PINEDA and ANDREAS G. ANDREOU, The Johns Hopkins University 70

7:30 IM:22 *A STUDY OF PARALLEL PERTURBATIVE GRADIENT DESCENT*
D. LIPPE and J. ALSPECTOR, Bellcore 70

7:30 IM:23 *IMPLEMENTATION OF NEURAL HARDWARE WITH THE NEURAL VLSI OF URAN IN APPLICATIONS OF REDUCED REPRESENTATIONS*
IL-SONG HAN and YOUNG-JAE CHOI, Korea Telecom Research Center and KI-CHUL KIM and HWANG-SOO LEE, Korea Advanced Institute of Science and Technology 71

7:30 IM:24 *SINGLE TRANSISTOR LEARNING SYNAPSES*
PAUL HASLER, CHRIS DIORIO, BRADLEY A. MINCH and CARVER MEAD, California Institute of Technology 71

LEARNING THEORY

7:30 LT:21 *LIMITS ON LEARNING MACHINE ACCURACY IMPOSED BY DATA QUALITY*
CORINNA CORTES, L.D. JACKEL and WAN-PING CHIANG, AT&T Bell Laboratories 71

7:30 LT:22 *LEARNING FROM QUERIES FOR MAXIMUM INFORMATION GAIN IN UNLEARNABLE PROBLEMS*
PETER SOLLICH and DAVID SAAD, University of Edinburgh 72

7:30 LT:23 *BIAS, VARIANCE AND THE COMBINATION OF LEAST SQUARES ESTIMATORS*
RONNY MEIR, Technion 72

7:30 LT:24 *ON-LINE LEARNING OF DICHOTOMIES*
N. BARKAI and H. SOMPOLINSKY, The Hebrew University and H.S. SEUNG, AT&T Bell Laboratories 72

7:30 LT:25 *DYNAMIC MODELLING OF CHAOTIC TIME SERIES WITH NEURAL NETWORKS*
JOSE C. PRINCIPE and JYH-MING KUO, University of Florida, Gainesville 73

7:30 LT:26 *A RIGOROUS ANALYSIS OF LINSKER'S HEBBIAN LEARNING NETWORK*
JIANFENG FENG, Universitat Tubingen, and HONG PAN and VWANI P. ROYCHOWDHURY, Purdue University 73

730 LT:27 *SAMPLE SIZE REQUIREMENTS FOR FEEDFORWARD NEURAL NETWORKS*
MICHAEL J. TURMON and TERRENCE L. FINE, Cornell University 73

7:30 LT:28 *ASYMPTOTICS OF GRADIENT-BASED NEURAL NETWORK TRAINING ALGORITHMS*
SAYANDEV MUKHERJEE and TERRENCE L. FINE, Cornell University 74

NEUROSCIENCE

7:30 NS:21 *SHORT-TERM ACTIVE MEMORY, INHIBITION, AND NEUROMODULATION: A COMPUTATIONAL MODEL OF PREFRONTAL CORTEX FUNCTION*
TODD S. BRAVER and JONATHAN D. COHEN, Cargegie Mellon University and DAVID SERVAN-SCHREIBER, University of Pittsburgh 74

7:30 NS:22 *A NEURAL MODEL OF DELUSIONS AND HALLUCINATIONS IN SCHIZOPHRENIA*
EYTAN RUPPIN and JAMES A. REGGIA, University of Maryland and DAVID HORN,Tel Aviv University 75

7:30 NS:23 *SPATIAL REPRESENTATIONS IN THE PARIETAL CORTEX MAY USE BASIS FUNCTIONS*
ALEXANDRE POUGET and TERRENCE J. SEJNOWSKI, The Salk Institute 75

7:30 NS:24 *GROUPING COMPONENTS OF THREE-DIMENSIONAL MOVING OBJECTS IN AREA MST OF VISUAL CORTEX*
RICHARD S. ZEMEL and TERRENCE J. SEJNOWSKI, The Salk Institute 76

7:30 NS:25 *A MODEL OF THE NEURAL BASIS OF THE RAT'S SENSE OF DIRECTION*
WILLIAM E. SKAGGS, JAMES J. KNIERIM, HEMANT S. KUDRIMOTI, and BRUCE L. MCNAUGHTON, University of Arizona, Tucson 76

SPEECH RECOGNITION

7:30 SP:21 *HIERARCHICAL MIXTURES OF EXPERTS APPLIED TO A FRAME-BASED NEURAL NETWORK SYSTEM FOR CONTINOUS SPEECH RECOGNITION*
YING ZHAO, RICHARD SCHWARTZ, and JOHN MAKHOUL, BBN System and Technologies 77

VISION

7:30 VI:21 *LEARNING DIRECTION IN GLOBAL MOTION: TWO CLASSES OF PSYCHOPHYSICALLY-MOTIVATED MODELS*
V. SUNDARESWARAN and LUCIA M. VAINA, Boston University 77

7:30 VI:22 *USING A NEURAL NET TO INSTANTIATE A DEFORMABLE MODEL*
CHRISTOPHER K.I. WILLIAMS, MICHAEL D. REVOW, and GEOFFREY E. HINTON, University of Toronto 77

7:30 VI:23 *DECORRELATION DYNAMICS: A THEORY FOR ORIENTATION CONTRAST AND ADAPTATION*
DAWEI W. DONG, University of California, Berkeley 78

7:30 VI:24 *NONLINEAR IMAGE INTERPOLATION USING SURFACE LEARNING*
CHRISTOPH BREGLER, University of California, Berkeley and STEPHEN M. OMOHUNDRO, Int. Computer Science Institute 78

7:30 VI:25 *COARSE-TO-FINE IMAGE SEARCH USING NEURAL NETWORKS*
CLAY D. SPENCE, JOHN C. PEARSON, and JIM BERGEN, David Sarnoff Research Center 78

THURSDAY AM

ORAL SESSION 9
ALGORITHMS & ARCHITECTURES

8:30 O9.1 *FINANCIAL APPLICATIONS OF LEARNING FROM HINTS (INVITED TALK)*
Y.S. ABU-MOSTAFA, California Institute of Technology 79

9:00 O9.2 *COMBINING ESTIMATORS USING NON-CONSTANT WEIGHTING FUNCTIONS*
VOLKER TRESP, Siemens AG, Central Research 79

9:20 O9.3 *AN INPUT OUTPUT HMM ARCHITECTURE*
YOSHUA BENGIO, Universite de Montreal and PAOLO FRASCONI, Universita di Firenze 79

9:40 O9.4 *BOLTZMANN CHAINS AND HIDDEN MARKOV MODELS*
LAWRENCE SAUL and MICHAEL JORDAN, MIT 79

10:00 *BREAK*

ORAL SESSION 10

ALGORITHMS & ARCHITECTURES

10:30 O10.1 *BAYESIAN QUERY CONSTRUCTION FOR NEURAL NETWORK MODELS*
GERHARD PAASS and JORG KINDERMANN, German National Research Center for Computer Science 80

10:50 O10.2 *USING A SALIENCY MAP FOR ACTIVE SPATIAL SELECTIVE ATTENTION:IMPLEMENTATION & INITIAL RESULTS*
SHUMEET BALUJA and DEAN A. POMERLEAU, Carnegie Mellon University 80

11:10 O10.3 *MULTIDIMENSIONAL SCALING AND DATA CLUSTERING*
THOMAS HOFMANN and JOACHIM BUHMANN, Rheinische Friedrich-Wilhelms-Universitat 80

11:30 O10.4 *A NON-LINEAR INFORMATION MAXIMISATION ALGORITHM THAT PERFORMS BLIND SEPARATION*
ANTHONY J. BELL and TERRENCE J. SEJNOWSKI, The Salk Institute 81

11:50 ADJOURN TO VAIL FOR WORKSHOPS

WORKSHOPS AT VAIL

DECEMBER 2, 1994

NOVEL CONTROL TECHNIQUES FROM BIOLOGICAL INSPIRATION

ORGANIZERS:Richard D. Braatz,(rdb@beethoven.che.caltech.edu), University of Illinois, James S. Schwaber,(schwaber@eplrx7.es.dupont.com), DuPont, David Touretzky,(dst@CS.CMU.EDU), Carnegie Mellon, Thomas F. Enders, Technical University Munich, K. P. Unnikrishnan,(unni@neuro.cs.gmr.com) General Motors

82

Panel participants: Martin Pottmann, DuPont, Babatunde A. Ogunnaike, DuPont, James Keeler, MCC, Austin, Michael A. Henson, Louisiana State University, Gerald Dreyfus, ESPCI, Paris, Francis J. Doyle, Purdue 82

MORNING SESSION:

7:30 Dave Touretzky and A. David Redish summarize their cognitive neuroscience theory of rodent navigation with implications for hippocampal function, and its implementation on a mobile robot.

82

8:00 discussion period for Touretsky/Redish presentation

82

8:15 Thomas F. Enders and collaborators summarize their research efforts in using neural networks in the development of techniques for the scheduling, control,and on-line optimization of batch fermentation processes (e.g. the alcoholic fermentation with yeast).

82

8:45 discussion period for Enders et al. presentation

83

9:00 panel/general discussion 83

AFTERNOON SESSION:

4:30 James S. Schwaber, Richard D. Braatz, Francis J. Doyle, Michael A. Henson, Martin Pottmann, and Babatunde A. Ogunnaike summarize their research efforts in developing novel process control techniques via inspiration from the cardiorespiratory reflexes.

83

5:00 discussion period for Schwaber et al. presentation

83

5:15 other workshop attendees present their work

83

6:00 panel/general discussion 83

83

MACHINE LEARNING APPROACHES IN COMPUTATIONAL MOLECULAR BIOLOGY

ORGANIZERS:Pierre Baldi (pfbaldi@juliet.caltech.edu),Soren Brunak (brunak@cbs.dth.dk)

83

MORNING SESSION:

7:30 Pierre Baldi, "Hidden Markov Models of Human Genes" 84

8:00 Soren Brunak, "Construction of Low Similarity Data Sets of Sequences with Funtional Sites for Prediction Purposes" 84

8:30 Tim Hunkapiller 84

9:00 Anders Krogh, "Predicting Protein Secondary Structure with Structured Networks" 84

AFTERNOON SESSION:

4:30 Paul Stolortz, "Links between statistical physics and dynamic programming:applications to computational molecular biology" 84

5:00 Gary Stormo, "Neural Networks for the Identification of Functional Domains Common to Multiple Sequences" 84

5:30 Niels Tolstrup, "Neural Network Model of the Genetic Code" 84

6:00 Discussion 84

NOVELTY DETECTION AND ADAPTIVE SYSTEM MONITORING

ORGANIZERS: Thomas Petsche (petsche@scr.siemens.com) and Stephen J. Hanson (jose@learning.siemens.com), Siemens Corporate Research, Inc.; Mark Gluck (gluck@pavlov.rutgers.edu), Rutgers University 84

7:30 - 9:00 Helicopter gearbox monitoring presentations and discussions by RobertR. Kolesar (ONR), Kourosh Danai (U Mass), Peter Kazlas (U Colorado, Boulder) and Mark Gluck (Rutgers). 85

4:30 - 6:00 Engine and electric motor monitoring by Ken Marko (Ford), Scott Smith (Boeing), and Thomas Petsche (Siemens). 85

Recognizing novelty in classification tasks by Germano Vasconcelos University of Kent) and Dimitrios Bairaktaris (University of Stirling). 85

85

ANTHROPOMORPHIC SPEECH SIGNAL PROCESSING

ORGANIZERS: Hynek Hermansky (hynek@eeap.ogi.edu) and Misha Pavel (pavel@eeap.ogi.edu) Oregon Graduate Institue 85

MORNING SESSION:

7:30 Jont Allen (Bell Laboratories, Murray Hill), "Speech Recognition with Human Face" 86

8:00 Andreou Andreas (Johns Hopkins University), "Analog Auditory Models" 86

8:30 Malcom Slaney (Interval Research), "Correlograms" 86

9:00 Discussion 86

AFTERNOON SESSION:

4:30 Nelson Morgan (International Computers Science Institute and U C Berkeley),"Current Research in Stochastic Perceptual Auditory-event-based Models (SPAM) " 86

5:00 Chalapathy Neti (IBM Watson Center), "Neuromorphic speech processing for speech recognition in noisy environments." 86

COMPUTATIONAL ROLE OF LATERAL CONNECTIONS IN THE CORTEX

ORGANIZER:Joseph Sirosh , UT Austin 86

MORNING SESSION:

7:30 Gary Blasdel: Title to be announced. 87

8:00 Terrence Sejnowski: "Physiological Effects of Intrinsic Horizontal Connections in Visual Cortex" 87

8:30 Jack Cowan: "Geometric Visual Hallucinations and Lateral Cortical Connections" 87

AFTERNOON SESSION:

4:30 Shimon Edelman: "Computational models of 3D object representation in the visual cortex, and the possible role of lateral connections" 87

5:00 Jonathan Marshall: "Do lateral connections help stabilize perception during occlusion events?" 87

5:30 DeLiang Wang: "Lateral connections and coherent oscillations" 87

6:00 Joseph Sirosh: "Cooperative self-organization of lateral connections and feature detectors in the visual cortex" 87

87

87

UNSUPERVISED LEARNING RULES AND VISUAL PROCESSING

ORGANIZERS:Lei Xu (lxu@cs.cuhk.hk) and Laiwan Chan (lwchan@cs.cuhk.hk), The Chinese University of Hong Kong; Zhaoping Li (lwchan@cs.cuhk.hk), Hong Kong University of Science and Technology 88

88

MORNING SESSION 1:Chair, Lei Xu

7:30 John Wyatt and Ibrahim Elfadel (MIT), "Time-Domain Solutions of Oja's Equations" 88

7:50 Leon Bottou (Neuristique Paris) and Yoshua Bengio (University of Montreal), "Kmeans Performs Newton Optimization" 88

8:10 Lei Xu (The Chinese University of Hong Kong and Peking University), "Multisets Modeling Learning: An Unified Framework for Unsupervised Learning" 88

8:30 Nathan Intrator (Tel-Aviv University), "Information Theory Motivation For Projection Pursuit" 88

9:00 Peter Dayan (University of Toronto),"The Helmholtz Machine" 88

EVENING SESSION 1:Chair, Zhaoping Li

4:30 Juergen Schmidhuber (Technische Universitaet Muenchen), "Predictability Minimization And Visual Processing" 89

4:50 Tony Bell (Salk Institute), "Non-linear, Non-gaussian Information Maximisation: Why It's More Useful" 89

5:10 Zhaoping Li (Hong Kong University of Science and Technology),"Understanding The Visual Cortical Coding From Visual Input Statistics" 89

5:30 Klaus Obermayer (Universitaet Bielefel), "Formation Of Orientation And Ocular Dominance In Macaque Striate Cortex" 89

5:50 Joseph Sirosh (University of Texas at Austin), "Putative Functional Roles Of Self-organized Lateral Connectivity In The Primary Visual Cortex" 89

6:00 Discussion 89

MORNING SESSION 2:Chair, Laiwan Chan

7:30 Yoshua Bengio (University of Montreal),"Density Estimation with a Hybrid of Neural Networks and Gaussian Mixtures" 89

7:50 Eric Mjolsness (UCSD) and Steve Gold (Yale University), "Learning Object Models through Domain-Specific Distance Measures" 89

8:10 Dit-Yan Yeung (Hong Kong University of Science and Technology),"Auto-associative Learning of On-line Handwriting Using Recurrent Neural Networks" 89

8:30 Volker Tresp (Siemens AG, Central Research), "Training Mixtures of Gaussians with Deficient Data" 89

8:50 George F. Harpur and Richard W. Prager (Cambridge University), "A Fast Method for Activating Competitive Self-Organizing Neural-Networks" 89

EVENING SESSION 2:Chair, Lei Xu

4:30 Michael E. Hasselmo (Harvard University), "Neuromodulatory Mechanisms For Regulation Of Cortical Self-organization" 89

4:50 Sue Becker (McMaster University), "Learning To Cluster Visual Scenes With Contextual Modulation" 89

5:10 Jonathan A. Marshall (University of North Carolina at Chapel Hill),"Invisibility in Vision: Occlusion, Motion, Grouping, and Self-Organization" 89

5:30 Irwin King and Lei Xu (The Chinese University of Hong Kong), "A Comparative Study on Receptive Filters by PCA Learning and Gabor Functions" 89

5:50 Bernd Fritzke (Ruhr-Universitaet Bochum), "Detection of Visual Feature Locations with a Growing Neural Gas Network" 89

6:10 Discussion 89

STATISTICAL AND NEURAL NETWORK APPROACHES TO NATURAL LANGUAGE PROCESSING

ORGANIZERS:Gary Cottrell (gary@cs.ucsd.edu) 90

90

FRIDAY MORNING:

90

7:30 AM Mitch Marcus: "Statistical approaches to NLP"
90

8:00 AM Gary Cottrell: "Neural net approaches to NLP"Learning fsa's and pda's
90

8:30 AM Lee Giles "Learning a class of large finite state machines with a recurrent neural network" 90

8:50 AM Sreerupa Das "Differentiable symbol processing and an application to language induction" 90

9:10 AM Patrick Juola and James Martin: "Extraction of Transfer Functions through Psycholinguistic Principles" 90

FRIDAY AFTERNOON:

4:30 PM George Berg "Single Network Approaches to Connectionist Parsing"
90

4:50 PM Ajay Jain, "PARSEC: Let Your Network do the Walking, but Tell it Where to Go."
90

5:10 PM Stan Kwasny: "Training SRNs to Learn Syntax"
90

5:30 PM Risto Miikkulainen "Parsing with modular networks"
90

5:50 PM - 6:30 PM The assembled crew 91

SATURDAY MORNING

7:30 AM Hinrich Schuetze: "Unsupervised word sense disambiguation for improved text retrieval"
91

7:50 AM David Yarowsky "A comparison of word sense disambiguation algorithms"
91

8:10 AM Nick Chater "Neural networks as statistical inference: Why it's best to have all one's assumptions out in the open" 91

8:30 AM Eric Brill "Statistical language processing: What are numbers good for?"
91

8:50 AM - 9:30 AM The assembled crew 91

SATURDAY AFTERNOON

4:30 PM Michael Gasser "Modular networks for language acquisition: Why and how"
91

4:50 PM David Plaut "Learning arbitrary and quasi-regular mappings in word reading with attractor networks" 91

5:10 PM Mark St. John "Practice makes perfect: The key role of construction frequency in sentence comprehension" 91

5:30 PM Kim Plunkett (unconfirmed), "Learning the Arabic plural: The case for minority default mappings in connectionist nets." 91

5:50 PM - 6:30 PM The assembled crew 91

NEURAL NETWORKS IN MEDICINE

ORGANIZER:Paul E. Keller (pe_keller@gate.pnl.gov)
91

92

FRIDAY MORNING:

7:30 AM Optimizing networks for Atlas guided segmentation of brain images, Anand Rangarajan, Yale University 92

8:00 AM Neural Net Analysis of Solitary Pulmonary Nodules, Armando Manduca, Mayo Clinic 92

8:30 AM Using Neural Networks for Semi-automated Pap Smear Screening, Laurie Mango, MD, and James M. Herriman, Neuromedical Systems Inc. 92

9:00 AM Automated design of optical-morphological structuring elements for Pap smear screening, J. P. Sharpe, R. Narayanswamy, N. Sungar*, H. Duke, R. J. Stewart, L. McKeogh and K. M. Johnson, University of Colorado at Boulder and *California Polytechnic State University 92

FRIDAY AFTERNOON:

4:30 PM Comparing the prediction accuracy of statistical models and artificial neural networks in breast cancer, Harry Burke, MD, David Rosen, Phil Goodman, MD, New York Medical University and University of Nevada 92

5:00 PM Diagnosis of hepatoma by committee, Bambang Parmanto and Paul Munro, University of Pittsburg 92

5:30 PM Discussion 92

92

SATURDAY MORNING:

7:30 AM Neural Networks for Nonlinear Processing of Biomagnetic/Bioelectric Signals, Martin Schlang, Michael Haft, and Ralph Neuneier, Siemens 93

8:00 AM Neural networks distinguish demented subjects from elderly controls based on EEGs, Beatrice Golomb, MD, and Andrew F. Leuchter, MD, UCLA 93

8:30 AM Normal and Abnormal EEG Classification using Neural Networks and other techniques, Ah Chung Tsoi, University of Queensland 93

9:00 AM Issues in Controlling Cardiac Chaos, Gary W. Flake, Siemens Corporate Research 93

SATURDAY AFTERNOON:

4:30 PM Prediction and Control of the Glucose Metabolism of a Diabetic, Volker Tresp, John Moody* and Wolf-Ridiger Delong, Siemens and *Oregon Graduate Institute 93

5:00 PM Experiences in using neural networks for detecting coronary artery disease, Georg Doffner, Austrian Institute of Artificial Intelligence - University of Vienna 93

5:30 Panel Discussion 93

ADVANCES IN RECURRENT NETWORKS

ORGANIZERHava Siegelmann (iehava@ie.technion.ac.il): 93

93

FRIDAY MORNING:

Lee A. Feldkamp (Remarks on Time-Lagged RNN--Training and Applications) 93

Jerry Connor (bootstrap methods in time series prediction) 93

Paul Muller (Programmable Aanlog Neural Computer: Design and Performance) 93

Lee Shung (Learning with smoothing Regularization) 93

Manuel Samuelides (application: design of neuro-filters) 93

94

Gary Kuhn (application of sensitivity analysis) 94

Morten With Pederson (Training and Pruning) 94

FRIDAY AFTERNOON:

Paolo Frasconi - Learning and Rule Embedding 94

Lei Xu - Mixture Models and the EM Algorithm 94

Hava Siegelmann - Towards a Neural Language: Symbolic to Analog 94

General discussion 94

SATURDAY MORNING

Pierre Baldi - Trajectory Learning Using Shallow Hierarchies of Oscillators 94

Mahesan Niranjan - Stacking Multiple RNN Models of the Vocal Tract 94

Kenji Doya - Problems Concerning Bifurcations of Network Dynamics 94

Hugo deGaris - The CAM-Brain Project : Evolution of a Billion Neuron Brain 94

Dawei Dong - Associative Dynaimc Decorrelation 94

SATURDAY AFTERNOON

Yoshua Bengio - On the Problem of Learning with Long-Term Dependencies 94

Barak Pearlmuter - On the Alleged Difficulty of Learning Long-Term Dependencies 94

Ricard Gavalda - On the Kolmogorov Complexity of RNN 94

Panel Discussion 94

DECEMBER 3, 1994

OPEN AND CLOSED PROBLEMS IN NEURAL NETWORK ROBOTICS

ORGANIZER: Marcus Mitchell (marcus@hope.caltech.edu)Chris M Bishop (Aston University)
95

95

MORNING SESSION:

7:30 - 7:35 Opening Remarks, Marcus Mitchell, Caltech
95

7:35 - 8:00 Why it's harder to control your robot than your arm: closed, open and irrelevant issues in inverse kinematics, Dave Demers, UCSD
95

8:05 - 8:30 Open Problem: Optimal Motor Hidden Units, Terry Sanger, JPL
95

8:35 - 9:00 Neural Network Vision for Outdoor Robot Navigation, Dean Pomerleau, CMU
95

AFTERNOON SESSION:

4:30 - 4:55 Learning New Representations and Strategies, Chris Atkeson, Georgia Tech
95

95

5:00 - 5:25 A Semi-Crisis for Neural Network Robotics:, Formal Specification of Robot Learning Tasks, Andrew Moore, CMU 96

5:30 - 6:30 Closing Discussion 96

NEURAL NETWORK ARCHITECTURES WITH TIME DELAY CONNECTIONS

ORGANIZERS:Andrew D. Back (back@elec.uq.oz.au), Eric A. Wan (ericwan@eeap.ogi.edu
96

96

MORNING SESSION:

7:30-7:45 Opening Discussion - Andrew Back, University of Queensland
96

7:45-8:00 "Computational Capabilities of Local-Feedback Recurrent Networks", Paolo Frasconi, University of Florence, Italy 96

8:00-8:15 " Issues in Representation: Recurrent Networks as Sequential Machines", C. Lee Giles and B.G. Horne, NEC Research Institute 96

8:15-8:30 "Properties of Recursive Memory Structures", Jose C. Principe, University of Florida
96

8:30-8:45 "A Local Model Net Approach to Modeling Nonlinear Dynamic Systems", Roderick Murray-Smith. MIT 96

8:45-9:15 Open forum: 5 minute presentations by participants
96

9:15-9:30 Question Time and Discussion 96

AFTERNOON SESSION:

4:30-4:45 "A Spatio-Temporal Approach to Visual Pattern Recognition", Lokendra Shastri, ICSI 96

4:45-5:00 "The Performance of Recurrent Networks for Classifying Time-Varying Patterns", Tina Burrows and Mahesan Niranjan, Cambridge University Engineering Department 96

5:00-5:15 "Nonlinear Infomax With Adaptive Time Delays", Tony Bell, The Salk Institute 97

5:15-5:30 "The Sinc Tensor Product Network", Jerome Soller, University of Utah 97

5:30-5:45, "Discriminating Between Mental Tasks Using a Variety of EEG Representations", Chuck Anderson, Colorado State University 97

5:45-6:00 Open forum: 5 minute presentations by participants 97

6:00-6:30 Question Time and Closing Discussion 97

97

ALGORITHMS FOR HIGH DIMENSIONAL SPACES: WHAT WORKS AND WHY

ORGANIZER: MICHAEL P. PERRONE, (mpp@watson.ibm.com) 97

97

MORNING SESSION:

7:30 "Statistical Properties of High Dimensional Spaces", Michael Perrone (IBM T.J. Watson Research Center) 97

8:00 "Computational Learning and Statistical Prediction", Jerome Friedman (Stanford University) 97

8:30 "Discriminant Adaptive Nearest Neighbor Classification", Trevor Hastie and Rob Tibshirani (Stanford University) 98

9:00 "Local Methods in High Dimension: Are They Surprisingly Good But Miscalibrated?", David Rosen (New York Medical College) 98

AFTERNOON SESSION:

4:30 "Is There Anything Positive in High Dimensional Spaces?", Nathan Intrator (Tel Aviv University) 98

5:00 "Three Techniques for Dimension Reduction", John Moody (Oregon Graduate Institute) 98

5:30 "A Local Linear Algorithm for Fast Dimension Reduction", Nandakishore Kambhatla (Oregon Graduate Institute) 98

6:00 "Fuzzy Dimensionality Reduction", Yinghua Lin (Los Alamos National Lab) 98

DOING IT BACKWARDS: NEURAL NETWORKS AND THE SOLUTION OF INVERSE PROBLEMS

ORGANIZER: Chris M Bishop (Aston University) 99

99

MORNING SESSION:

7:30 "Welcome and overview" Chris Bishop (Aston) 99

7:35 "From ill-posed problems to all neural networks and beyond through regularization" Tomaso Poggio / Federico Girosi (MIT) 99

7:55 "Solving inverse problems using an EM approach to density estimation" Zoubin Ghahramani (MIT) 99

8:15 "Density estimation with periodic variables" Chris Bishop (Aston) 99

8:35 "Doing it forwards, undoing it backwards: high-dimensional compression and expansion" Russell Beale (University of Birmingham) 99

8:55 "Inversion of feed-forward networks by gradient descent" Alexander Linden (Berkeley) 99

9.15 Discussion 99

AFTERNOON SESSION:

4:30 "An iterative inverse of a talking machine" Sid Fels (Toronto) 99

4:50 "Diagnostic problem solving" Sungzoon Cho (Postech, S Korea) 99

5:10 "Multiple Models in Inverse Filtering of the Vocal Tract" M Niranjan (Cambridge) 99

5:30 "Goal directed model inversion" Silvano Colombano (NASA Ames) 99

5:50 "Predicting element concentrations in the SSME exhaust plume" Kevin Whitaker (University of Alabama) 99

6:10 Discussion 100

THE NEURAL BASIS OF LOCOMOTION: MODELS OF PATTERN GENERATORS

ORGANIZER: BARD ERMENTROUT 100

TUTORIAL PROGRAM

November 28, 1994

Session I: 9:30-11:30 am

9:30-11:30 **Recent Advances in Learning Theory**
Michael Kearns, AT&T Bell Laboratories

This tutorial will provide a detailed survey of some of the new models and most significant results in computational learning theory over the past several years. Special emphasis will be given to methods of analysis and algorithmic techniques that have the potential to be useful beyond just the formal settings considered in the literature so far. No prior knowledge of computational learning theory is required.

In the first part of the tutorial, issues of computational efficiency will be central as we survey the known results on learning three powerful classes of representations- neural networks, disjunctive normal form expressions, and decision trees. These three classes are now almost completely understood in several learning models, and we will examine several elegant algorithms based on the powerful techniques of Fourier analysis over a basis of parity functions

In the second part of the tutorial, we concentrate on statistical or information-theoretic issues. We will describe recent analyses of Bayesian weighting schemes that provide worst-case mistake bounds for on-line prediction without requiring any underlying assumptions on the data. We will also survey learning curve behavior such as sudden drops in generalization error. If time permits, applications of this latter topic to problems of structural risk minimization will be discussed.

Michael Kearns received a Ph.D. in computer science from Harvard University in 1989. Following postdoctoral fellowships at MIT and the International Computer Science Institute, Kearns joined the research staff at AT&T Bell Laboratories, where he has been since 1991. With Umesh Vazirani of U.C. Berkeley, he has recently completed "An Introduction to Computational Learning Theory", which is being published by The MIT Press in August of 1994.

9:30-11:30 **A Survey of Pattern Recognition Hardware**
Dan Hammerstrom, OGI and Adaptive Solutions Inc.

This tutorial will look at a variety of hardware devices designed for applications that loosely fall under the topic of pattern recognition. The orientation of this tutorial will be towards building systems with commercial applicability. The hardware studied will include examples of specialized architectures for neural network emulation such as the AT&T ANNA, Bellcore Boltzmann Engine, Adaptive Solutions CNAPS, and Intel/Nestor Ni1000, as well as systems designed for image and vision processing, such as the TI MVP and Martin Marietta GAPP. We will also study the basic motivation for creating such hardware and the cost/performance justification in the face of competition from high performance RISC and DSP engines.

We will also discuss the deadly effects of Amdahl's Law and the flexibility versus performance/cost trade-offs in the VLSI design space.

Dan Hammerstrom received a B.S. degree from Montana State University, a M.S. degree from Stanford University, and the Ph.D. degree from the University of Illinois, all in Electrical Engineering. He was on the faculty of Cornell University from 1977 to 1980 as an Assistant Professor. From 1980 to 1985 he worked for Intel where he participated in the development and implementation of the iAPX-432 and i960, and, as a consultant, on the iWarp systolic processor. He is founder and Chief Technical Officer of Adaptive Solutions, Inc., and is also an Associate Professor at the Oregon Graduate Institute. He has been a Visiting Professor at the Royal Institute of Technology in Stockholm, Sweden. Dr. Hammerstrom's research interests are in the area of the VLSI implementation of neural network structures. He has been an Associate Editor for the Journal of the International Neural Network Society, IEEE Transactions on Neural Networks, and the International Journal of Neural Networks.

Session II: 1:00-3:00 pm

1:00-3:00 **Advances in the Theory and Applications of the Self-Organizing Map**
Teuvo Kohonen, Helsinki University of Technology

The basic Self-Organizing Map (SOM) is a computational algorithm that places a number of "codebook" vectors (parameter vectors) into the space of input signals in an ordered fashion. The process in which the codebook vectors are determined may be characterized as a kind of nonparametric regression.

.The most important and typical applications of the SOM are visualization of complex data, automatic discovery of abstract relations from raw data, and adaptive control of robots and processes. Hundreds of different applications of the SOM have already been developed.

This tutorial contains the following topics: Introduction to neural computing and competitive learning in particular. The basic SOM algorithms. The SOM for generalized distance metric. Dynamically defined neighborhood functions. Operator maps. Batch computation of the SOM. Semantic SOMs. Applications. Physiological interpretation of the SOM. Learning Vector Quantization(LVQ) and its applications. No specific prerequisites for this tutorial are necessary.

Teuvo Kohonen, Dr. Eng., is a Professor of Computer Science at the Helsinki University of Technology, Finland, and Permanent Research Professor of the Academy of Finland. His research areas are associative memories, neural networks, and pattern recognition, in which he has published over 200 research papers and three monography books. His fourth book is on digital computers. Since the 1960's, Professor Kohonen has introduced several new concepts to neural computing: fundamental theories of distributed associative memory and optimal associative mappings, the learning subspace method, the self-organizing feature maps, the learning vector quantization, and novel algorithms for symbol processing like the redundant hash addressing and dynamically expanding context. The best known application of his work is the neural speech recognition system.

1:00-3:00 Learning to Act: An Introduction to Reinforcement Learning

Andy Barto, University of Massachusetts at Amherst

There is increasing interest in the reinforcement learning paradigm because it addresses problems faced by autonomous agents. In this tutorial, I describe the contributions of a number of researchers who are treating reinforcement learning as a collection of methods for successively approximating solutions to stochastic optimal control problems. Within this framework, methods for improving heuristic evaluation functions by "backing up" evaluations can be understood in terms of dynamic programming solutions to optimal control problems. Such methods include one used by Samuel in his checkers playing program of the late 1950's, Holland's Bucket-Brigade algorithm, connectionist Adaptive Critic methods, Watkins' Q-Learning, and Korf's Learning-Real-Time-A* algorithm. Establishing the connection between evaluation function learning and the extensive theory of optimal control and dynamic programming produces a number of immediate results as well as a sound theoretical basis for future research.

Andrew G. Barto is a Professor of Computer Science, University of Massachusetts, Amherst. He received a B.S. with distinction in mathematics, 1970, and a Ph.D. in Computer Science, 1975, University of Michigan. Core faculty of the Neuroscience and Behavior Program, University of Massachusetts. He is a member of the Society for Neuroscience, INNS, Cognitive Science Society, is a senior member of the IEEE, and a member and Fellow of the American Association for the Advancement of Science. Professor Barto was elected to the INNS board of governors 1991. He is an associate editor for *Neural Computation*, member of the editorial board for *Neural Networks*, action editor for *Machine Learning*, and an associate editor for the MIT Press book series *Neural Network Modeling and Connectionism*. Professor Barto's research centers on learning in natural and artificial systems, and he has studied learning algorithms for artificial neural networks since 1977, contributing to the development of *associative reinforcement learning* methods and their application to control problems. Current research centers on models of motor learning and learning mehtods for real-time planning and control.

Session III: 3:30-5:30 pm

3:30-5:30 Images of the Mind: A Tutorial on Brain Imaging

Marc Raichle, Washington University Medical School (St. Louis)

It has been suggested repeatedly that the human brain actually possesses two means of generating response: a non-automatic, attention-demanding mechanism and an effortless, automatic mechanism. The existence of two pathways for verbal response selection, one for novel tasks and one for learned tasks, provides the framework for efficient operation based upon the development of what some would call habits. What has been missing is a m,eans by which we could test such a hypothesis in terms of the underlying normal human brain circuitry.

The introduction of modern brain imaging techniques such as x-ray computed tomography (CT), positron emission tomography (PET), and, more recently, magnetic resonance imaging (MRI) has changed the situation dramatically. It is now possible to examine safely not only the physical anatomy of the living brain with x-ray, CT, and MRI, but also its functional anatomy (i.e., the actual areas involved in specific tasks) with PET and, more recently, with MRI. We can now visualize the brain circuitry involved in a variety of cognitive tasks including verbal response selection and determine the validity of the hypothesis that two circuits underlie verbal response selection. From this rapidly evolving brain imaging literature emerges not only strong support for the two route hypothesis but also a clear indication of the actual anatomical circuits involved.

Professor Raichle did some of the seminal work in PET imaging helping to usher in the modern era of functional anatomy and imaging. Professor Raichle just published with M. Posner the recent Scientific American volume "Images of Mind".

3:30-5:30 Statistics and Nets: Understanding Nonlinear Models from Their Linear Relatives

Leo Breiman, University of California at Berkeley

Linear regression is a good testbed for many important issues regarding general regression problems. Uning linear regressions to study these issues is analogous to testing new treatments on mice. They have a simple structure, and compute very fast. This makes theoretical investigation and extensive simulations possible. The result is that many interesting questions have been extensively studied in the linear regression context. A good deal of this work has implications for general nonlinear regression problems.

In this tutorial, I will summarize work that has wider applications. This includes issues like variable selection versus methods that shrink coefficents instead of zeroing them; accuracy of leave-one-out cross-validation vs. leave many out; stacked regressions; prediction of multiple correlated responses; and the effects of instability on prediction accuracy. I will also give a brief review of the ideas and measures of influence of data points on parameter estimates.

Leo Breiman is Professor of Statistics and Director of the Statistical Computing Facility at the University of California at Berkeley. He is co-author of "Classification and Regression Trees" (CART), and author of three other books on probability and statistics. His field of research is in nonlinear methods for classification and regression.

TUESDAY AM
ORAL SESSION 1

COGNITIVE NEUROSCIENCE

8:30 **O1.1 THE PROBLEM OF VISUAL AWARENESS (INVITED TALK)**
F.H. CRICK, Salk Institute.

9:00 **O1.2 DIRECTION SELECTIVITY IN PRIMARY VISUAL CORTEX USING MASSIVE INTRACORTICAL CONNECTIONS**
HUMBERT SUAREZ, Caltech, CHRISTOF KOCH, Caltech, and RODNEY DOUGLAS, University of Oxford.

Almost all models of orientation and direction selectivity in visual cortex are based on feed-forward connection schemes, where geniculate input provides all the excitation to both pyramidal and inhibitory neurons. The latter neurons then suppress the response of the former for non-optimal stimuli. However, anatomical studies show that up to 90% of the excitatory synaptic input onto any cortical cell is provided by other cortical cells. The massive excitatory feedback nature of cortical circuits is embedded in the *canonical microcircuit* of Douglas & Martin (1991). We here investigate analytically and through biologically realistic simulations the functioning of a detailed model of this circuitry. In the model, weak geniculate input is dramatically amplified by the action of intracortical excitation, while inhibition has a dual role: (i) to prevent the early geniculate-induced excitation in the null direction and (ii) to restrain excitation and ensure that the neurons fire only when the stimulus is in their receptive-field. Among the insights gained are the possibility that hysteresis underlies visual cortical function, paralleling proposals for short-term memory, and strong limitations on linearity tests using gratings for cortical neurons. We compare in detail properties of visual cortical neurons to this model and to a classical model of direction selectivity that does not include excitatory cortico-cortical connections. The model explain a number of puzzling features about direction selective simple cells, including the small somatic input conductance changes that have been measured experimentally during stimulation in the null direction. The model also allows us to understand why the velocity-response curve of area 17 neurons is different from that of their LGN afferents, and the origin of expansive and compressive nonlinearities in the contrast-response curve of striate cortical neurons.

9:20 **SPOTLIGHT I: COGNITIVE NEUROSCIENCE**

PLASTICITY AS LIKELIHOOD OF RELEVANCE: COMPETITION IN DISTRIBUTED REPRESENTATIONS, Nicol N. Schraudolph and Terrence J. Sejnowski, Computational Neurobiology Laboratory, Salk Institute

GRAMMAR LEARNING BY A SELF-ORGANIZING NETWORK, Michiro Negishi, Boston University

PATTERNS OF DAMAGE IN NEURAL NETWORKS: THE EFFECTS OF LESION AREA, SHAPE AND NUMBER,Eytan Ruppin and James A Reggia, University of Maryland

9:30 **O1.3 ON THE COMPUTATIONAL UTILITY OF CONSCIOUSNESS**

DONALD W. MATHIS AND MICHAEL C. MOZER, University of Colorado

We propose a computational framework for understanding and modeling human consciousness. This framework integrates many existing theoretical perspectives, yet is sufficiently concrete to allow simulation experiments. We do not attempt to explain qualia (subjective experience and feelings), but instead ask what differences exist within the cognitive information processing system when a person is conscious of some information versus when that information is unconscious. The central idea we explore is that the contents of consciousness correspond to temporally stable states in an interconnected network of specialized computational modules. Each module functions as an associative memory that operates in two stages: (1) a fast, essentially feedforward, input-output mapping that attempts to achieve an appropriate response to a given input, and (2) a slower relaxation search that is concerned with achieving semantically well-formed states. It is the stable attractors of the relaxation search that reach conscious awareness. To illustrate the operation of a module, we model performance on a simple arithmetic task and show that the sequence of stable states in our model corresponds roughly to the conscious mental states people experience when performing this task. What might be the computational utility of stable states within the cognitive architecture? Our simulations show that periodically settling to stable states improves performance by cleaning up inaccuracies and noise, forcing decisions, and helping to keep the system on track toward a solution.

9:50 **O1.4 TEMPORAL CHARACTERISTICS OF DYNAMIC MOTOR LEARNING**

TOM BRASHERS-KRUG, EMANUEL V. TODOROV, and REZA SHADMEHR, MIT

Biological sensorimotor systems are not static maps that transform input (sensory information) into output (motor behavior). Evidence from many lines of research suggests that the representations in this system are plastic, experience-dependent entities. If the sensorimotor system configures itself to perform well under one set of circumstances, will it then necessarily perform poorly when placed in an environment with radically different demands? Or can it learn and retain two anti-correlated mappings? We present psychophysical and computational results that explore this question in the context of a dynamic motor learning task. We find that, although all subjects demonstrate temporal crosstalk when learning two negatively correlated dynamic environments, some are still able to segregate the differing demands of the tasks and to form and maintain two incompatible input\output mappings. Modular neural networks are well suited for the demands of this task. By adding a simple temporal component to the grating units of such a network, we were able to account for the more unexpected aspects of some subjects' behavior.

10:10 **BREAK**

ORAL SESSION 2

REINFORCEMENT LEARNING

10:40 **O2.1 REINFORCEMENT LEARNING ALGORITHM FOR PARTIALLY OBSERVABLE MARKOV DECISION PROBLEMS**

TOMMI JAAKKOLA, SATINDER P. SINGH, and MICHAEL I. JORDAN, MIT

Increasing attention has been paid to reinforcement learning algorithms in recent years, partly due to successes in the theoretical analysis of their behavior in Markov environments. If the Markov assumption is removed, however, neither the algorithms nor the analyses continue to be usable. We propose and analyze a new learning algorithm to solve a certain class of non-Markov decision problems. Our algorithm applies to problems in which the environment is Markov, but the learner has restricted access to state information. The algorithm involves a Monte-Carlo policy evaluation combined with a policy improvement method that is similar to that of Markov decision problems and is guaranteed to converge to a local maximum. The algorithm operates in the space of stochastic policies, a space which can yield a policy that performs considerably better than any deterministic policy. Although the space of stochastic policies is continuous-even for a discrete action space-our algorithm is computationally tractable. Computationally tractable policy improvement schemes for such continuous spaces have not been available previously. Finally, we note that the algorithm neatly combines exploration with policy refinement.

11:00 **O2.2 ADVANTAGE UPDATING APPLIED TO A DIFFERENTIAL GAME**

MANCE E. HARMON, LEEMON C. BAIRD III, and A. HARRY KLOPF, Wright Laboratory

An application of reinforcement learning to a linear-quadratic, differential game is presented. The reinforcement learning system uses a recently developed algorithm, the residual gradient form of advantage updating. The game is a Markov Dicision Process (MDP) with continuous time, states, and actions, linear dynamics, and a quadratic cost function. The game consists of two players, a missile and a plane; the missile pursues the plane and the plane evades the missile. The reinforcement learning algorithm for optimal control is modified for differential games in order to find the minimax point, rather than the maximum. Simulation results are compared to the optimal solution, demonstrating that the simulated reinforcement learning system converges to the optimal answer. The performance of both the residual gradient and non-residual gradient forms of advantage updating and Q-learning are compared. The results show that advantage updating converges faster than Q-learning in all simulations. The results also show advantage updating converges regardless of the time step duration; Q-learning is unable to converge as the time step duration grows small.

11:20 **SPOTLIGHT II: REINFORCEMENT LEARNING**

OPTIMAL MOVEMENT PRIMITIVES,Terence Sanger, Jet Propulsion Laboratory

AN INTEGRATED ARCHITECTURE OF ADAPTIVE NEURAL NETWORK CONTROL FOR DYNAMIC SYSTEMS, Liu Ke, Robert L. Tokar,and Brian D. McVey, Los Alamos National Laboratory

PHASE-SPACE LEARNING, Fu-Sheng Tsung and Garrison W. Cottrell, University of California, San Diego

11:30 **02.3 REINFORCEMENT LEARNING WITH SOFT STATE AGGREGATION**

SATINDER P. SINGH, TOMMI JAAKKOLA, and MICHAEL I. JORDAN, MIT

It is widely accepted that the use of more compact representations than lookup tables is crucial to scaling reinforcement learning (RL) algorithms to real-world problems. Unfortunately almost all of the theory of reinforcement learning (RL) assumes lookup table representations. In this paper we address the pressing issue of combining function approximation and RL, and present 1) a function approximator based on a simple extension to state aggregation (a commonly used form of compact representation), namely *soft* state aggregation, 2) a theory of convergence for RL with arbitrary, but fixed, soft state aggregation, 3) a novel intuitive understanding of the effect of state aggregation on online RL, and 4) a new heuristic *adaptive* state aggregation algorithm that finds improved compact representations by exploiting the non-discrete nature of soft state aggregations. Preliminary empirical results are also presented.

11:50 **02.4 GENERALIZATION IN REINFORCEMENT LEARNING: SAFELY APPROXIMATING THE VALUE FUNCTION**

JUSTIN A. BOYAN and ANDREW W. MOORE, Carnegie Mellon University

A straightforward approach to the curse of dimensionality in reinforcement learning and dynamic programming is to replace the lookup table with a generalizing function approximator such as a neural net. Although this has been successful in the domain of backgammon, there is no guarantee of convergence. In this paper, we show that the combination of dynamic programming and function approximation is not robust, and in even very benign cases, may diverge and produce an entirely wrong policy. We then introduce Grow-Support, a new algorithm which is safe from divergence yet can still reap the benefits of successful generalization.

12:00 **LUNCH**

Preprint jab@cs.cmu.edu

TUESDAY PM
ORAL SESSION 3

NEUROSCIENCE

2:00 **O3.1 SEEING AND DECIDING: A WINNER-TAKE-ALL DECISION PROCESS IN THE CEREBRAL CORTEX(INVITED TALK)**
W.T. NEWSOME, Stanford University School of Medicine

2:30 **O3.2 A MODEL FOR CHEMOSENSORY RECEPTION**
RAINER MALAKA and THOMAS RAGG, Universitat Karlsruhe, and MARTIN HAMMER, Freie Universitat Berlin

A new model for chemosensory reception is presented. It models reactions between odor molecules and receptor proteins and the activation of second messenger by receptor proteins. The mathematical formulation of the reaction kinetics is transformed into an artificial neural network (ANN). The resulting feed-forward network provides a powerful means for parameter fitting by applying learning algorithms. The weights of the network corresponding to chemical parameters can be trained by presenting experimental data. We demonstrate the simulation capabilities of the model with experimental data from honey bee chemosensory neurons. It can be shown that our model is sufficient to rebuild the observed data and that simpler models are not able to do this task.

2:50 **SPOTLIGHT III: NEUROSCIENCE**
MODEL OF A BIOLOGICAL NEURON AS A TEMPORAL NEURAL NETWORK, Sean D. Murphy and Edward W. Kairiss, Yale University

A CRITICAL COMPARISON OF MODELS FOR ORIENTATION AND OCULAR DOMINANCE COLUMNS IN THE STRIATE CORTEX, Ed Erwin and Klaus Obermayer, Universitat Bielefeld

A NOVEL REINFORCEMENT MODEL OF BIRDSONG VOCALIZATION LEARNING, Kenji Doya and Terrence J. Sejnsowski, Howard Hughes Medical Institute, Salk Institute

3:00 **O3.3 THE ELECTRONIC TRANSFORMATION: A TOOL FOR RELATING NEURONAL FORM TO FUNCTION**
NICHOLAS T. CARNEVALE, KENNETH Y. TSAI, AND THOMAS H. BROWN, Yale University and BRENDA J. CLAIBORNE, University of Texas

The spatial distribution and time course of electrical signals in neurons have important theoretical and practical consequences. Because it is difficult to infer how neuronal form affects electrical signaling, we have developed a quantitative yet intuitive approach to the analysis of electrotonus. This approach transforms the architecture of the cell from anatomical to electrotonic space, using the logarithm of voltage attenuation as the distance metric. We describe the theory behind this approach and illustrate its use.

3:20 **FEEDBACK REGULATION OF CHOLINERGIC MODULATION AND AUTO-ASSOCIATIVE MEMORY FUNCTION IN HIPPOCAMPAL REGION CA3**

MICHAEL E. HASSELMO, EDI BARKAI and JOSHUA BERKE, Harvard University

Models of the hippocampus have proposed autoassociative memory function in region CA3 and heteroassociative memory function at the connections from CA3 to CA1. However, they have not considered how the recall of previously stored patterns might interfere with the learning of new information. Acetylcholine may set the appropriate dynamics for learning within the hippocampal formation. A model of the feedback regulation of cholinergic modulation in hippocampal region CA3 was developed, focusing on the putative autoassociative function of synapses arising from CA3 pyramidal cells. Feedback regulation of cholinergic modulation allowed the network to respond to novel patterns with strong cholinergic modulation, allowing accurate learning, and to respond to familiar patterns with a decrease in cholinergic modulation, allowing recall. This function required that suppression of synaptic transmission in region CA3 and CA1 be stronger for synapses in stratum radiatum (arising from region CA3), in contrast to synapses in stratum lacunosum-moleculare (arising from the entorhinal cortex). Experiments in brain slice preparations of the hippocampus demonstrate this laminar selectivity.

3:40 **BREAK**

ORAL SESSION 4

LEARNING THEORY

Neuropulse maass, spiking, ps, z

4:15 **O4.1 ON THE COMPUTATIONAL COMPLEXITY OF NETWORKS OF SPIKING NEURONS**

WOLFGANG MAASS, Technische Universitaet Graz, Austria

We investigate the computational power of a formal model for networks of spiking neurons, and provide bounds for the number of examples that are needed to train such networks.

4:35 **O4.2 H^{∞} OPTIMAL TRAINING ALGORITHMS AND THEIR RELATION TO BACKPROPAGATION**

BABAK HASSIBI and THOMAS KAILATH, Stanford University

We derive *global* H^{∞} optimal training algorithms for neural networks. These algorithms guarantee the smallest possible prediction error energy over all possible disturbances of fixed energy, and are therefore robust with respect to model uncertainties and lack of statistical information on the exogenous signals. The ensuing estimators are infinite-dimensional, in the sense that updating the weight vector estimate requires knowledge of all previous weight estimates. A certain finite-dimensional approximation to these estimators is the backpropagation algorithm. This explains the *local* H^{∞} optimality of backpropagation that has been previously demonstrated.

4:55 **SPOTLIGHT IV: LEARNING THEORY**

RESPONSE FUNCTIONS FOR LEARNING IN LARGE LINEAR PERCEPTRONS, Peter Sollich, University of Edinburgh

GENERALISATION IN FEEDFORWARD NETWORKS, Adam Kowalczyk and Herman Ferra, Telecom Australia, Research Laboratories

FROM DATA DISTRIBUTIONS TO REGULARIZATION IN INVARIANT LEARNING, Todd K. Leen, Oregon Graduate Institute of Science and Technology

NEURAL NETWORK ENSEMBLES, CROSS VALIDATION, AND ACTIVE LEARNING, Anders Krogh and Jesper Vedelsby, Technical University of Denmark

5:10 **O4.3 SYNCHRONY AND DESYNCHRONY IN OSCILLATOR NETWORKS**

DE LIANG WANG AND DAVID TERMAN,Ohio State University

An novel class of locally excitatory, globally inhibitory oscillator networks is proposed and investigated analytically and by computer simulation. The model of each oscillator corresponds to a standard relaxation oscillator with two time scales. The network exhibits a mechanism of selective gating, whereby an oscillator jumping up to its active phase rapidly recruits the oscillators stimulated by the same pattern, while preventing other oscillators from jumping up. We show analytically that with the selective gating mechanism the network rapidly achieves both synchronization within blocks of oscillators that are stimulated by connected regions and desynchronization between different blocks. Computer simulations demonstrate the network's promising ability for segmenting multiple input patterns in real time. This model lays a physical foundation for the oscillatory correlation theory of feature binding, and may provide an effective computational framework for pattern segmentation and figure/ground segregation.

5:30 **DINNER**

7:30 **REFRESHMENTS AND POSTER SESSION I**

TUESDAY EVENING POSTERS

ALGORITHMS & ARCHITECTURES

7:30 **AA:1 EXTRACTING RULES FROM ARTIFICIAL NEURAL NETWORKS WITH DISTRIBUTED REPRESENTATIONS**

SEBASTIAN THRUN, University of Bonn

Although artificial neural networks have been applied with remarkable success to a variety of real-world scenarios, they have often been criticized for exhibiting a low degree of human comprehensibility. Techniques that compile compact sets of symbolic rules out of artificial neural networks offer a promising perspective to overcome this obvious shortcoming of neural network representations.

This paper presents an approach to the extraction of *if-then* from artificial neural networks. Its key mechanism is *validity interval analysis*, which is a generic tool for extracting symbolic knowledge by propagating rule-like knowledge through Backpropagation-style neural networks. Empirical studies in a robot arm kinematics domain illustrate the appropriateness of the method to extract rules from networks with real-valued and distributed representations.

7:30 **AA:2 CAPACITY AND INFORMATION EFFICIENCY OF A BRAIN-LIKE ASSOCIATIVE NET**

BRUCE GRAHAM and DAVID WILLSHAW

We have determined the capacity and information efficiency of an associative net configured in a brain-like way with partial connectivity and noisy input cues. Recall theory was used to calculate the capacity when pattern recall is achieved using a *winners-take-all* strategy. Transforming the *dendritic sum* according to *input activity* and *unit usage* can greatly increase the capacity of the associative net under these conditions. Maximum information efficiency was achieved with very low connectivity levels ($\leq$10%). This corresponds to the level of connectivity commonly seen in the brain and invites speculation that the brain is connected in the most information efficient way.

7:30 AA:3 BOOSTING THE PERFORMANCE OF RBF NETWORKS WITH DYNAMIC DECAY ADJUSTMENT

MICHAEL R. BERTHOLD, Forschungszentrum Informatik and JAY DIAMOND, Intel Corp.

Radial Basis Function (RBF) Networks, also known as networks of locally-tuned processing units are well known for their ease of use. Most algorithms used to train these types of networks, however, require a fixed architecture, in which the number of units in the hidden layer must be determined before training starts. The RCE training algorithm, introduced by Reilly, Cooper and Elbaumand its probabilistic extension the P-RCE algorithm, take advantage of a growing structure in which hidden units are only introduced when necessary. The nature of these algorithms allows training to reach stability much faster than is the case for gradient-descent based methods. Unfortunately P-RCE networks do not adjust the standard deviation of their prototypes, using only one global value for this parameter. This paper introduces the Dynamic Decay Adjustment (DDA) algorithm which utilizes the *constructive* nature of the P-RCE algorithm together with independent adaptation of each prototype's decay factor. In addition, this radial adjustment is class dependent and distinguishes between different neighbours. It is shown that networks trained with the presented algorithm perform substantially better than common RBF networks.

7:30 AA:4 SIMPLIFYING NETWORKS BY DISCOVERING "FLAT" MINIMA

SEPP HOCHREITER and JURGEN SCHMIDHUBER, Technische Universitat Munchen

We present a new algorithm for finding low complexity networks with high generalization capability. The algorithm searches for large connected regions of so-called "flat" minima of the error function. In the weight-space environment of a "flat" minimum, the error remains (approximately) constant. Our method is theoretically justified: under a broad and reasonable range of conditions weaker than those used in previous work, it can be shown that flat minima correspond to low expected overfitting. Although our algorithm requires the computation of second order derivatives, it has the same order of complexity as backprop. In experiments with feedforward and recurrent nets, the method clearly outperforms conventional gradient descent, by finding networks with minimal complexity and theoretically optimal generalization performance.

7:30 AA:5 LEARNING WITH PRODUCT UNITS

LAURENS R. LEERINK and MARWAN A. JABRI, University of Sydney, and C. LEE GILES and BILL G. HORNE, NEC Research Institute

Product units provide a method of automatically learning the higher-order input combinations required for efficient learning in neural networks. However, we show that problems are encountered when using backpropagation to train networks containing these units. This paper examines these problems, and proposes some atypical heuristics to improve learning. Using these heuristics a constructive method is introduced which solves well-researched problems with significantly less neurons than previously reported. Secondly, product units are implemented as candidate units in the Cascade Correlation (Fahlman & Lebiere, 1990) system. This resulted in smaller networks which trained faster than when using sigmoidal or Gaussian units.

7:30 **AA:6 DETERMINISTIC ANNEALING VARIANT OF THE EM ALGORITHM**

NAONORI UEDA and RYOHEI NAKANO, NTT Communication Science Laboratories

We present a deterministic annealing variant of the EM algorithm for maximum likelihood parameter estimation problems. In our approach, the EM process is reformulated as the minimization problem of a thermodynamic free energy by using the *principle of maximum entropy* and *statistical mechanics analogy*. Unlike simulated annealing approach, the minimization is *deterministically* performed. Moreover, the deried algorithm, unlike the conventional EM algorithm, can obtain better estimates free of the initial parameter values.

7:30 **AA:7 PLASTICITY AS LIKELIHOOD OF RELEVANCE: COMPETITION IN DISTRIBUTED REPRESENTATIONS**

NICOL N. SCHRAUDOLPH and TERRENCE J. SEJNOWSKI, Salk Institute

Conventionally, differentiation within groups of redundant sub-units in a neural network is achieved through competition for activity. Simple inhibition is limited to local representations, while decorrelation and factorization schemes that support distributed representations are computationally unattractive. We introduce an alternative framework in which competition is mediated by neural *plasticity* instead, leading to diffuse, nonadaptive competitive mechanisms for distributed representations. We show how this versatile approach can improve an unsupervised feature detection rule we have previously proposed, as well as greatly increase the speed of convergence for deep backpropagation networks.

7:30 **AA:8 DIFFUSION OF CREDIT IN MARKOVIAN MODELS**

YOSHUA BENGIO, Universite de Montreal and PAOLO FRASCONI, Universita de Firenze, Italy

This paper studies the problem of diffusion in Markovian models (such as hidden Markov models) and how it makes very difficult the task of learning of long-term dependencies in sequences.

7:30 **AA:9 MIXTURE OF ONE-DIMENSIONAL PROJECTIONS (MODP): A UNIFYING ARCHITECTURE FOR PRINCIPAL COMPONENT ANALYSIS AND COMPETITIVE LEARNING**

JOSHUA B. TENENBAUM AND EMANUEL V. TODOROV, MIT

We present a unifying framework for two basic styles of unsupervised learning, *Principal Components Analysis (PCA)* and *Competitive Learning (CL)*. In this framework, the transition from PCA to CL is an example of the tradeoff between model simplicity and model accuracy. We present a single flexible architecture for unsupervised learning, the *Mixture of One-Dimensional Projections (MODP)*, that optimizes this tradeoff. MODP moves smoothly between PCA and CL modes by adjusting the relative weights of the complexity cost and the reconstruction cost. We show that MODP supports a spectrum of robust representations between strict PCA and strict CL, and that it is capable of adapting its style of representation to the character of a particular data set.

7:30 AA:10 INTERIOR POINT IMPLEMENTATIONS OF ALTERNATING MINIMIZATION TRAINING

MICHAEL LEMMON and PETER T. SZYMANSKI, University of Notre Dame

This paper presents an alternating minimization training algorithm used in the training of radial basis function networks. The algorithm is a modification of an interior point method used in solving primal linear programs. The resulting algorithm is shown to have a convergence rate on the order of $\sqrt{n}L$ iterations where n is a measure of the network size and L is a measure of the resulting solution's accuracy

7:30 AA:11 SARDNET: A SELF-ORGANIZING FEATURE MAP FOR SEQUENCES

DANIEL L. JAMES and RISTO MIIKKULAINEN, University of Texas at Austin

A self-organizing neural network for sequence classification called SARDNET is described and analyzed experimentally. SARDNET extends the Kohonen Feature Map architecture with activation retention and decay in order to create unique distributed response patterns for differnt sequences. SARDNET yields extremely dense yet descriptive representations of sequential input in very few training iterations. The network has proven successful on mapping arbitrary sequences of binary and real numbers, as well as phonemic representations of English words. Potential applications include isolated spoken word recognition and cognitive science models of sequence processing.

7:30 AA:12 CONVERGENCE PROPERTIES OF THE K-MEANS ALGORITHMS

LEON BOTTOU, Neuristique and YOSHUA BENGIO, Universite de Montreal

This paper studies the convergence properties of the well known K-Means clustering algorithm. The K-Means algorithm can be described either as a gradient descent algorithm or by slightly extending the mathematics of the EM algorithm to this hard threshold case. We show that the K-Means algorithm actually minimizes the quantization error using the very fast Newton algorithm.

7:30 AA:13 ACTIVE LEARNING FOR FUNCTION APPROXIMATION

KAH KAY SUNG and PARTHA NIYOGI, MIT

We develop a principled strategy to sample a function optimally for function approximation tasks within a Bayesian framework. Using ideas from optimal experiment design, we introduce a novel objective function to measure the degree of approximation, and the potential utility of the data points towards optimizing this objective. We show how the general strategy can be used to derive precise algorithms to select data for two cases: learning unit step functions and polynomial functions. In particular, we investigate whether such active algorithms can learn the target with fewer examples. We obtain theoretical and empirical results to suggest that this is the case.

7:30 AA:14 PHASE-SPACE LEARNING

FU-SHENG TSUNG and GARRISON W. COTTRELL, University of California, San Diego

Existing recurrent net learning algorithms are inadequate. We introduce the conceptual framework of viewing recurrent training as matching vector fields of dynamical systems in phase space. Phase-space reconstruction techniques make the hidden states explicit, reducing temporal learning to a feed-forward problem. In short, we propose viewing *iterated prediction* (Lapedes & Farber, 1988) as the best way of training recurrent networks on deterministic signals. Using this framework, we can train multiple trajectories, insure their stability, and *design* arbitrary dynamical systems.

7:30 AA:15 ANALYSIS OF UNSTANDARDIZED CONTRIBUTIONS IN CROSS CONNECTED NETWORKS

THOMAS R. SHULTZ, YURIKO OSHIMA-TAKANE, and YOSHIO TAKANE, McGill University

Understanding knowledge representations in neural nets has been a difficult problem. Principal components analysis (PCA) of contributions (products of sending activations and connection weights) have yielded valuable insights into knowledge representations, but much of this work has focused on the correlation matrix of contributions. The present work shows that analyzing the variance-covariance matrix of contributions yields more valid insights.

7:30 AA:16 TEMPLATE-BASED ALGORITHMS FOR CONNECTIONIST RULE EXTRACTION

JAY A. ALEXANDER and MICHAEL C. MOZER, University of Colorado

Casting neural network weights in symbolic terms is crucial for interpreting and explaining the behavior of a network. Additionally, in some domains, a symbolic description could lead to more robust generalization. We present a principled approach to symbolic rule extraction based onthe notion of *weight templates*, parameterized regions of weight space corresponding to specific symbolic expressions. With an appropriate choice of representation, we show how template parameters may be efficiently identified and instantiated to yield the optimal match to a unit's actual weights. Depending on the requirements of the application domain, our method can accommodate n-ary disjunctions and conjunctions with O(k) complexity, simple n-of-m expressions with $O(k^2)$ complexity, or a more general class or recursive n-of-m expressions with $O(k^3)$ complexity, where k is the number of inputs to a unit. Our method of rule extraction offers several benefits over alternative approaches in the literature, and simulation results on a variety of problems demonstrate its effectiveness.

COGNITIVE SCIENCE

7:30 CS:1 GRAMMAR LEARNING BY A SELF-ORGANIZING NETWORK

MICHIRO NEGISHI, Boston University

This paper presents the design and simulation results of a self-organizing neural network which induces grammar from example sentences. Input sentences are generated from a simple phase structure grammar including recursive noun phrase construction rules. The network induces a grammar explicitly in the form of symbol categorization rules and phrase production rules.

7:30 CS:2 FORWARD DYNAMIC MODELS IN HUMAN MOTOR CONTROL: PSYCHOPHYSICAL EVIDENCE

DANIEL M. WOLPERT, ZOUBIN GHAHRAMANI and MICHAEL I. JORDAN, MIT

The notion of an internal model, a system which mimics the behavior of a natural process, has emerged as an important theoretical concept in motor control. Based on purely engineering principles, with as yet no experimental validation, it has been recently proposed that the central nervous system (CNS) uses a forward dynamic model of its motor plant in planning, control and learning. We have used a novel experimental approach, based on investigating the propagation of errors in state estimation, to assess both the existence of a forward dynamic model of the arm and its nature. We present both experimental results and simulations which, taken together, support the existence of such a model.

7:30 CS:3 PATTERNS OF DAMAGE IN NEURAL NETWORKS: THE EFFECTS OF LESION AREA, SHAPE AND NUMBER

EYTAN RUPPIN and JAMES A. REGGIA, University of Maryland

This paper presents a general analytical framework for estimating the functional damage resulting from focal structural lesions to a neural network. This framework is used to study theoretically the effects of focal lesions of varying area, shape and number on the retrieval capacities of a spatially-organized associative memory. Although our analytical results are based on some approximations, they correspond well with simulation results. Our study sheds light on some important features characterizing the clinical manifestations of multi-infarct dementia, including its classical step-wise progression, the strong association between the number of infarcts and the prevalence of dementia after stroke, and the 'multiplicative' interaction that has been postulated to occur between Alzheimer's disease and multi-infarct dementia.

CONTROL

7:30 CN:1 OPTIMAL MOVEMENT PRIMITIVES

TERENCE D. SANGER, Jet Propulsion Laboratory

The theory of Optimal Unsupervised Motor Learning shows how a network can discover a reduced-order controller for an unknown nonlinear system by representing only the most significant modes. Here, I extend the theory to apply to command sequences, so that the most significant components discovered by the network correspond to motion "primitives". Combinations of these primitives can be used to produce a wide variety of different movements. I demonstrate applications to human handwriting decomposition and synthesis, as well as to the analysis of electrophysiological experiments on movements resulting from stimulation of the frog spinal cord.

7:30 **CN:2 AN INTEGRATED ARCHITECTURE OF ADAPTIVE NEURAL NETWORK CONTROL FOR DYNAMIC SYSTEMS**

LIU KE, ROBERT L. TOKAR, and BRIAN D. McVEY, Los Alamos National Laboratory

In this study, an integrated neural network control architecture for nonlinear dynamic systems is presented. Most of the recent emphasis in the neural network control field has no error feedback as the control input, which rises the lack of adaptation problem. The integrated architecture in this paper combines feed forward control and error feedback adaptive control using neural networks. The paper reveals the different internal functionality of these two kinds of neural network controllers for certain input styles, e.g., state feedback and error feedback. Feed forward neural network controllers with state feedback establish fixed control mappings which can not adapt when model uncertainties are present. With error feedback, neural network controllers learn the slopes or the gains with respect to the error feedback, producing an error driven adaptive control system. The results demonstrate that the two kinds of control scheme can be combined to realize their individual advantages. Testing with disturbances added to the plant shows good tracking and adaptation with the integrated neural control architecture.

IMPLEMENTATIONS

7:30 **IM:1 PULSESTREAM SYNAPSES WITH NON-VOLATILE ANALOGUE AMORPHOUS-SILICON MEMORIES**

A.J. HOLMES, A.F. MURRAY, S. CHURCHER and J. HAJTO, University of Edinburgh, and M.J. ROSE, Dundee University

A novel two-terminal device, consisting of a thin 1000 A layer of $p^{\prime}$ a-Si:H sandwiched between Vanadium and Chromium electrodes, exhibits a non-volatile, analogue memory action. A circuit has been designed in which this device stores synaptic weights in an ANN chip, replacing the capacitor previously used for dynamic weight storage. Two different synapse designs are discussed and results are presented.

7:30 **IM:2 A LAGRANGIAN FORMULATION FOR TRAINING OF KERR-TYPE OPTICAL NETWORKS**

JAMES E. STECK, STEVEN R. SKINNER, and ELIZABETH C. BEHRMAN, The Wichita State University

A training method based on a form of continuous spatially distributed optical error backpropagation is presented for an optical network composed of nondiscrete neurons and weighted interconnections. The optical network is feed-forward and is composed of thin layers of a Kerr-type self-focusing/defocusing nonlinear optical material. The training method is derived from a Lagrangian formulation of the constrained minimization of the network error at the output. This leads to a formulation that describes training as a calculation of the distributed error of the optical signal at the output which is then reflected back through the device to assign a spatially distributed error to the internal layers. This error is then used to modify the internal weighting values. Results from several computer simulations of the training are presented, and a simple optical table demonstration of the training is discussed.

7:30 IM:3 A CHARGE-BASED CMOS PARALLEL ANALOG VECTOR QUANTIZER

GERT CAUWENBERGHS, John Hopkins University and VOLNEI PEDRONI, California Institute of Technology

We present an analog VLSI chip for parallel analog vector quantization The MOSIS 2.0 μ m double-poly CMOS Tiny chip contains an array of 16 x 16 charge-based distance estimation cells, implementing a mean absolute difference (MAD) metric operating on a 16-input analog vector field and 16 analog template vectors. The distance cell including dynamic template storage measures 60 x 78 μm^2. Additionally, the chip features a winner-take-all (WTA) output circuit of linear complexity, with global positive feedback for fast and decisive settling of a single winner output. Experimental results on the complete 16 x 16 VQ system demonstrate correct operation with 34 dB analog input dynamic range and 3 μ sec cycle time at 0.7 mW power dissipation.

7:30 IM:4 AN AUDITORY LOCALIZATION AND COORDINATE TRANSFORM CHIP

TIMOTHY HORIUCHI, California Institute of Technology

The localization and orientation to various novel or interesting events in the environment is a critical sensorimotor ability in all animals, predator or prey. In mammals, the superior colliculus (SC) plays a major role in this behavior, the deeper layers exhibiting responses to visual, auditory, and somatosensory stimuli. While the different sensory modalities are naturally in different coordinates, the representation in the SC is found to be retinotopic. Auditory cues, in particular, are thought to be computed in head-based coordinates which must be transformed to retinal coordinates. In this paper, an analog VLSI implementation for auditory localization is described which extends the barn owl architecture to primates where further transformation is required due to moveable eyes. This transformation is intended to model the projection in primates from auditory cortical areas to the deeper layers of the primate superior colliculus. This system was developed to interface with an analog VLSI-based saccadic eye movement system also being constructed in our laboratory.

LEARNING THEORY

7:30 LT:1 HIGHER ORDER STATISTICAL DECORRELATION WITHOUT INFORMATION LOSS

GUSTAVO DECO, Siemens, AG, and WILFRIED BRAUER, Technische Universitat Munchen

A neural network learning paradigm based on information theory is proposed as a way to perform in an unsupervised fashion, redundancy reduction among the elements of the output layer without loss of information from the sensory input. The model developed performs nonlinear decorrelation up to higher orders of the cumulant tensors and results in probabilistically independent components of the output layer. This means that we don't need to assume Gaussian distribution neither at the input nor at the output. The theory presented is related to the unsupervised-learning theory of Barlow, which proposes redundance reduction as the goal of cognition. When nonlinear units are used (sigmoid or higher order pi-neurons) nonlinear principal component analysis is obtained. In this case nonlinear manifords can be reduced to minimum dimension manifolds. If such units are used the network performs a generalized principal component analysis in the sense that non-Gaussian distributions can be linearly decorrelated and higher orders of the correlation tensors are also taken into account. The basic structure of the architecture involves a general transformation that conserves the volume and therefore the entropy, yielding a map without loss of information. Minimization of the mutual information among the output neurons eliminates the redundancy between the outputs and results in statistical decorrelation of the extracted features. This is known as factorial learning. To sum up, this paper presents a model of factorial learning for general nonlinear transformations of an arbitrary non-Gaussian (or Gaussian) environment with statistically non-linearly correlated input. Simulations demonstrate the effectiveness of this method.

7:30 LT:2 HYPERPARAMETERS, EVIDENCE AND GENERALISATION IN AN UNREALISABLE SCENARIO

GLENN MARION and DAVID SAAD, University of Edinburgh

Using a statistical mechanical formalism we calculate the evidence, the generalisation error and the consistency measure for a linear perceptron trained on a set of examples generated by a nonlinear teacher. Our model allows us to interpolate between the known linear case and an unrealisable and nonlinear case. A comparison of the hyperparameters which maximixe the evidence with those that optimise the other performance measures reveals that, in the unrealisable case, the evidence procedure is a highly misleading guide to optimising the performance measures considered.

7:30 LT:3 RESPONSE FUNCTIONS FOR LEARNING IN LARGE LINEAR PERCEPTRONS

PETER SOLLICH, University of Edinburgh

We present a new method for obtaining the response function G and its average G from which most of the properties of learning and generalization in linear perceptrons can be derived. We first rederive the known results for the limit of perceptrons of infinite size N and show explicitly that G is self-averaging in this limit. We then discuss extensions of our method to more general learning scenarios with anisotropic teacher space priors, input distributions, and weight decay terms. Finally, we use our method to calculate the finite N corrections of order $1/N$ to G and discuss the corresponding finite size effects on generalization and learning dynamics.

7:30 LT:4 GENERALIZATION DYNAMICS IN NEURAL NETWORKS

CHANGFENG WANG and SANTOSH S. VENKATESH, University of Pennsylvania

In this paper we extend previous results on the generalization dynamics of linear machines to general nonlinear machines. The major result is an exact characterization of how a given learning machine generalizes during the training process when it is trained with a learning algorithm based on minimization of the empirical error (or a modification of the empirical error). This explicit characterization fully determines the local behavior of algorithms in the vicinity of solutions. In particular, we are enabled to analytically determine an *optimal stopping time* when training should be stopped to achieve optimal generalization. The results can be interpreted in terms of a *time-dependent, effective machine size* which forms the link between generalization error and machine complexity during learning viewed as an evolving process in time. The different roles played by the complexity of the machine *class* and the complexity of the *specific* machine in the class during learning are also precisely demarcated.

The methodology adopted can be readily adapted to study the dynamical effect of regularization on the learning process. In this framework we are enabled to compare the relative benefits of regularization and early (optimal) stopping. Indeed, we show that the two approaches are similar, yet distinct, in nature and effect, each having advantages over the other. The analysis also provides guidelines for the selection of penalty functions for the regularization method and provides general (nontraditional) guidelines for machine size selection.

Since the generalization error is defined in terms of an abstract loss function, the results find wide applicability including but not limited to regression (square-error loss function) and density estimation (log-likelihood loss) problems. The basic theoretical tools involved are uniform convergence of probability measures in empirical process (VC-method) and Von Mises' method.

7:30 LT:5 STOCHASTIC DYNAMICS OF THREE-STATE NEURAL NETWORKS

TORU OHIRA, Sony Computer Science Laboratory and JACK D. COWAN, University of Chicago

We present here an analysis of the stochastic neurodynamics of a neural network composed of three-state neurons described by a master equation. An outer-product representation of the master equation is employed. In this representation, an extension of the analysis from two to three-state neurons is easily performed. We apply this formalism with approximation schemes to a simple three-state network and compare the results with Monte Carlo simulations.

7:30 LT:6 LEARNING STOCHASTIC PERCEPTRONS UNDER K-BLOCKING DISTRIBUTIONS

MARIO MARCHAND and SAEED HADJIFARADJI, University of Ottawa

We present a statistical method that PAC learns the class of stochastic perceptrons with arbitrary monotonic activation function and weights $w_i \in \{-1,0,+1\}$ when the probability distribution that generates the input examples is member of a family that we call *k-blocking distributions*. Such distributions represent an important step beyond the case where each input variable is statistically independent since the 2k-blocking family contains all the Markov distributions of order k. By stochastic perceptron we mean a perceptron which, upon presentation of input vector x, outputs 1 with probability $f\left(\sum_i w_i x_i - \theta\right)$. Because the same algorithm works for any monotonic (nondecreasing or nonincreasing) activation function f on Boolean domain, it handles the well studied cases of sigmoids and the "usual" radial basis functions.

7:30 LT:7 GENERALISATION IN FEEDFORWARD NETWORKS

ADAM KOWALCZYK, Telecom Australia Research Laboratories.

We discuss a model of consistent learning with an additional restriction on the probability distribution of training samples and the target concept. We show that the model provides a significant improvement on the upper bounds of sample complexity, i.e. the minimal number of random training samples allowing a selection of the hypothesis with a predefined accuracy and confidence. Further, we show that the model has the potential for providing a finite sample complexity even in the case of infinite VC-dimension as well as for a sample complexity below VC-dimension. This is achieved by linking sample complexity to a sort of average number of implementable dichotomies ofa training sample rather than the maximal size of a shattered sample, i.e. VC-dimension.

7:30 LT:8 FROM DATA DISTRIBUTIONS TO REGULARIZATION IN INVARIANT LEARNING

TODD K. LEEN, Oregon Graduate Institute of Science and Technology

Ideally pattern recognition machines provide constant output when the inputs are transformed under a group G of desired invariances. These invariances can be achieved by enhancing the training data to include examples of inputs transformed by elements of G, while leaving the corresponding targets unchanged. Alternatively the cost function for training can include a regularization term that penalizes changes in the output when the input is transformed under the group.

This paper relates the two approaches, showing precisely the sense in which the regularized cost function approximates the result of adding transformed examples to the training data. We introduce the notion of a probability distribution over the group transformations and use this to rewrite the cost function for the enhanced training data. We show that this is equivalent to the sum of the original cost function plus a regularizer. For unbiased models, the regularizer reduces to the intuitively obvious choice. For infinitesimal transformations the coefficient of the regularization term reduces to the variance of the distortions introduced into the training data, thus providing a simple bridge between the two approaches.s

7:30 LT:9 NEURAL NETWORK ENSEMBLES, CROSS VALIDATION, AND ACTIVE LEARNING

ANDERS KROGH and JESPER VEDELSBY, Technical University of Denmark

Learning of continuous valued functions using neural network ensembles (committees) can give improved accuracy, reliable estimation of the generalization error, and active learning. The *ambiguity* is defined as the variation of the output of ensemble members averaged over *unlabeled* data, so it quantifies the disagreement among the networks. We show how to use the ambiguity in combination with cross validation to give a reliable estimate of the ensemble generalization error, and how this type of ensemble cross validation can actually improve performance. By a generalization of query by committee, it is shown how the ambiguity can by used to select new training data to be labeled in an active learning scheme. Finally, it is shown how to estimate the optimal weights of the ensemble members using *unlabeled data*.

NEUROSCIENCE

7:30 NS:1 OCULAR DOMINANCE AND ACTIVATION DYNAMICS IN A UNIFIED SELF-ORGANIZING MODEL OF THE VISUAL CORTEX

JOSEPH SIROSH and RISTO MIIKKULAINEN, University of Texas at Austin

A neural network model for the self-organization of lateral connections and ocular dominance columns from uncorrelated binocular input is presented. The self-organizing process results in a network where (1) afferent weights of each neuron organize into smooth hill-shaped receptive fields primarily on one of the retinas, (2) neurons with common eye preference form connected, intertwined patches, and (3) lateral connections primarily link regions of the same eye preference. Similar self-organization of cortical structures has been observed experimentally in strabismic kittens. The lateral connectivity in the model mediates activation dynamics that could explain, for example, why divergent squinters cannot form a single, binocular percept even for a completely binocular input. The model provides a basis for computational study of cortical self-organization and plasticity, as well as dynamic perceptual processes such as feature grouping and binding.

7:30 NS:2 ANATOMICAL ORIGIN AND COMPUTATIONAL ROLE OF DIVERSITY IN THE RESPONSE PROPERTIES OF CORTICAL NEURONS

KALANIT GRILL SPECTOR, SHIMON EDELMAN and RAPHAEL MALACH, The Weizmann Institute of Science

The maximization of diversity of neuronal response properties has been recently suggested as an organizing principle for the formation of such prominent features of the functional architecture of the brain as the cortical columns and the associated patchy projection patterns (Malach, TINS 17:101, 1994). We report a computational study of two aspects of this hypothesis. First, we show that maximal diversity is attained when the ratio of dendritic and axonal arbor sizes is equal to one, as it has been found in many cortical areas and across species (Lund et al, Cerebral Cortex 3:148, 1993). Second, we show that maximization of diversity leads to better performance in two case studies: in systems of receptive fields implementing steerable/shiftable filters, and in matching spatially distributed signals, a problem that arises in visual tasks such as stereopsis, motion processing, and recognition.

7:30 NS:3 MODEL OF A BIOLOGICAL NEURON AS A TEMPORAL NEURAL NETWORK

SEAN D. MURPHY and EDWARD W. KAIRISS, Yale University

A biological neuron can be viewed as a device that maps a multidimensional temporal event signal (dendritic postsynaptic activations) into a unidimensional temporal event signal (action potentials). We have designed a network, the Spatio-Temporal Event Mapping (STEM) architecture, which can learn to perform this mapping for arbitrary biophysical models of neurons. Such a network appropriately trained, called a STEM cell, can be used in place of a conventional compartmental model in simulations where only the transfer function is important, such as network simulations. The STEM cell offers advantages over compartmental models in terms of computational efficiency, analytical tractability, and as a framework for VLSI implementations of biological neurons.

7:30 NS:4 A CRITICAL COMPARISON OF MODELS FOR ORIENTATION AND OCULAR DOMINANCE COLUMNS IN THE STRIATE CORTEX

ED ERWIN and KLAUS SCHULTEN, University of Illinois and KLAUS OBERMAYER, Universitat Bielefeld

More than ten of the most prominent models for the structure and for the activity dependent formation of orientation and ocular dominance columns in the striate cortex have been evaluated. We implemented those models on parallel machines, we extensively explored parameter space, and we quantitatively compared model predictions with experimental data which were recorded optically from macaque striate cortex.

In our contribution we present a summary of our results to date. Briefly, we find that (*i*) despite apparent differences, many models are based on similar principles and, consequently, make similar predictions. (*ii*) certain "pattern models" as well as the developmental "correlation-based learning" models disagree with the experimental data, and (*iii*) of the models we have investigated, "competitive Hebbian" models and the recent model of Swindale provide the best match with experimental data.

7:30 NS:5 A NOVEL REINFORCEMENT MODEL OF BIRDSONG VOCALIZATION LEARNING

KENJI DOYA and TERRENCE J. SEJNOWSKI, Salk Institute

In songbirds that learn to imitate a tutor, the auditory learning phase can be separate from the motor learning phase. We have developed a theoretical framework for song learning that accounts for response properties of neurons that have been observed in many of the nuclei that are involved in song learning. Specifically, we suggest that the *anterior forebrain pathway*, which is not needed for song production in the adult but is essential for song acquisition, provides weight perturbations and an evaluation of the resulting song for vocalization learning. A computer model was constructed for the motor learning phase that could replicate a real zebra finch song with 90% accuracy based on a spectrographic measure. The second generation of the model bird could replicate the tutor song with 96% accuracy.

7:30 NS:6 REINFORCEMENT LEARNING PREDICTS THE SITE OF PLASTICITY FOR AUDITORY REMAPPING IN THE BARN OWL

ALEXANDRE POUGET, CEDRIC DEFFAYET, and TERRENCE J. SEJNOWSKI, Salk Institute

The auditory system of the barn owl contains several spatial maps. In young barn owls raised with optical prisms over their eyes, these auditory maps are shifted to stay in register with the visual map. This suggests that the visual input imposes a frame of reference on the auditory maps. However, the optic tectum, the first site of convergence of visual with auditory information, is not the site of plasticity for the shift of the auditory maps; the plasticity occurs instead in the inferior colliculus, which contains an auditory map and projects into the optic tectum. We explored the possibility that learning in the inferior colliculus is mediated by a global reinforcement signal whose delivery is controlled by the foveal representation of the visual system. We found that a simple hebb rule gated by reinforcement can learn to appropriately adjust auditory maps. In addition, we show that reinforcement learning preferentially adjusts the weights n the inferior colliculus, as in the owl brain, even when the weights are free to change anywhere in the network. This observation raises the possibility that the site of learning does not have to be genetically specified, but could be the results of how the learning procedure interacts with the network architecture.

7:30 NS:7 MORPHOGENESIS OF THE LATERAL GENICULATE NUCLEUS: HOW SINGULARITIES AFFECT GLOBAL STRUCTURE

SVILEN TZONEV and KLAUS SCHULTEN, Beckman Institute, University of Illinois and JOSEPH G. MALPELI, University of Illinois

The macaque lateral geniculate nucleus (LGN) exhibits an intricate lamination pattern, which changes midway through the nucleus at a point coincident with small gaps due to the blind spot in the retina. We present a three-dimensional model of morphogenesis in which local cell interactions cause a wave of development of neuronal receptive fields to propagate through the nucleus and establish different lamination patterns. We examine the interactions between the wave and the localized singularities due to the gaps, and find that the gaps induce the change in lamination pattern. We explore critical factors in determining general LGN organization.

REINFORCEMENT LEARNING

7:30 RL:1 INSTANCE-BASED STATE IDENTIFICATION FOR REINFORCEMENT LEARNING

R. ANDREW McCALLUM, University of Rochester

When a robot's next course of action depends on information hidden from the sensors because of problems such as occlusion, restricted range, bounded field of view and limited attention, we say the agent suffers from the *hidden state problem*. State identification techniques use history information to uncover hidden state.

This paper presents *instance-based state identification*, a new approach to reinforcement learning with state identification. Noting that learning with history and learning in continuous spaces both share the property that they begin without knowing the granularity of the state space, the approach applies instance-based (or "memory-based") learning to history sequences-instead of recording instances in a continous geometrical space, we record instances in action-perception-reward sequence space.

The first implementation of this approach, called *Nearest Sequence Memory*, is simplistic, but it nonetheless learns with an order of magnitude fewer steps than several previous approaches.

7:30 RL:2 FINDING STRUCTURE IN REINFORCEMENT LEARNING

SEBASTIAN THRUN, University of Bonn and ANTON SCHWARTZ, Stanford University

Reinforcement learning addresses the problem of learning to select actions in order to maximize one's performance in unknown environments. It has been argued that in order to scale reinforcement learning to complex real-world tasks, such as typically studied in AI, one must ultimately be able to abstract away the myriad of details in the real world, and to operate in more appropriate and more tractable abstract problem spaces.

This paper presents the SKILLROY algorithm. SKILLROY discovers skills, which are macro-like action patterns that arise in the context of multiple, related tasks. Skills collapse whole action sequences into single operators. They are learned by minimizing the compactness of action policies, using a description length argument on their representation. Initial empirical results in simple grid navigation tasks are presented.

7:30 RL:3 REINFORCEMENT LEARNING METHODS FOR CONTINUOUS-TIME MARKOV DECISION PROBLEMS

STEVEN J. BRADTKE and MICHAEL O. DUFF, University of Massachusetts

Semi-Markov Decision Problems are continuous time generalizations of discrete time Markov Decision Problems. A number of reinforcement learning algorithms have been developed recently for the solution of Markov Decision Problems, based on the ideas of asynchronous dynamic programming and stochastic approximation. Among these are TD (λ) , Q-learning, and Real-time Dynamic Programming. After reviewing semi-Markov Decision Problems and Bellman's optimality equation in that context, we propose algorithms similar to those named above, adapted to the solution of semi-Markov Decision Problems. We demonstrate these algorithms by applying them to the problem of determining the optimal control for a simple queueing system. We conclude with a discussion of circumstances under which these algorithms may be usefully applied.

7:30 RL:4 A CLASS OF ACTOR/CRITIC ARCHITECTURES THAT ARE EQUIVALENT TO Q-LEARNING

ROBERT H. CRITES and ANDREW G. BARTO, University of Massachusetts

We prove the convergence of a class of actor/critic algorithms that are equivalent to Q-learning by construction. Their equivalence is achieved by encoding Q-values within the policy and value function of the actor and critic. The resultant actor/critic algorithms are novel in two ways: they update the critic only when the most probable action is executed from any given state, and they update the actor using criteria that depend on the relative probabilities of the actions that are taken.

SPEECH AND SIGNAL PROCESSING

7:30 SP:1 CONNECTIONIST SPEAKER NORMALIZATION WITH GENERALIZED RESOURCE ALLOCATING NETWORKS

CESARE FURLANELLO, DIEGO GIULIANI, and EDMONDO TRENTIN, Istituto per La Ricerca Scientifica e Tecnologica

Inter-speaker variability is one of the principal error sources in real-time automatic speech recognition. This paper presents a rapid speaker-normalization technique based on neural network spectral mapping. The neural network is used as a front end of a continuous speech recognition system (speaker-dependent. HMM-based) to normalize the input acoustic data from a new speaker. The difference between speakers can be reduced using a limited amount of new acoustic data (from 15 to 40 phonetically rich sentences). Recognition error of phone units from the acoustic-phonetic continuous speech corpus APASCI is decreased with an adaptability ratio of 42%. We used local basis network of elliptical gaussian kernels, with recursive allocation of units and on-line optimization of parameters (GRAN model). For this application, the model included a linear term. The results compare favorably with multivariate linear mapping based on constrained orthonormal transformations.

7:30 SP:2 USING VOICE TRANSFORMATIONS TO CREATE ADDITIONAL TRAINING SPEAKERS FOR WORD SPOTTING

ERIC I. CHANG and RICHARD P. LIPPMANN, MIT Lincoln Laboratory

Speech recognizers provide good performance for most users but the error rate often increases dramatically for a small percentage of talkers who are "different" from those talkers used for training. One expensive solution to this problem is to gather more training data in an attempt to sample these outlier users. A second solution, explored in this paper, is to artificially enlarge the number of training talkers by transrforming the speech of existing training talkers. This approach is similar to enlarging the training set for OCR digit recognition by warping the training digit images, but is more difficult bacause continuous speech has a much larger number of dimensions (e.g. linguistic, phonetic, style, temporal, spectral) that differ across talkers. Initial experiments explored the use of simple linear spectral warping to enlarge a 24-talker training data base used for word spotting. Transforming the original training conversations successfully increased the average detection rate of keywords that are suitable for this transformation by more than 10 percentage points. The average detection rate over all words was increased by 3.5 percentage points (from 67.8% to 71.3%). More complex speech transformations are currently being explored.

7:30 SP:3 A COMPARISON OF DISCRETE-TIME OPERATOR MODELS FOR NONLINEAR SYSTEM IDENTIFICATION

ANDREW D. BACK and AH CHUNG TSOI, University of Queensland

We present a unifying view of discrete-time operator models used in the context of finite word length linear signal processing. Comparisons are made between the recently presented gamma operator model, and the delta and rho operator models for performing nonlinear system identification and prediction using neural networks. A new model based on an adaptive bilinear transformation which generalizes all of the above models is presented.

VISION

7:30 VI:1 JPMAX: LEARNING TO RECOGNIZE MOVING OBJECTS AS A MODEL-FITTING PROBLEM

SUZANNA BECKER, McMaster University

Unsupervised learning procedures have been successful at low-level feature extraction and preprocessing of raw sensor data. So far, however, they have had limited success in learning higher-order representations, e.g., of objects in visual images. One way to force a network to discover higher-order structure is to make constraining assumptions about the kind of structure present in the data and build these constraints into the learning procedure. A promising approach is to maximize some measure of agreement between the outputs of two groups of neurons which receive inputs physically separated in space, time or modality, as in (Becker and Hinton, 1992; Becker, 1993; de Sa, 1993). Using the same approach, a much simpler learning procedure is proposed here which discovers features in a single-layer network consisting of several populations of neurons, and can be applied to multi-layer networks trained one layer at a time. We derive two cost functions which depend on the joint distribution of the populations' activities. Selection of an appropriate prior is of central importance. When trained with this algorithm on raw image sequences of moving geometric objects (circles, squares and triangles), a two-layer network can learn to perform accurate position-invariant object classification.

7:30 VI:2 PCA-PYRAMIDS FOR IMAGE COMPRESSION

HORST BISCHOF and KURT HORNIK, Technical University Vienna

This paper presents a new method for image compression by neural networks. The contribution of this paper is twofold. First, we show that we can use neural networks in a pyramidal framework, yielding the so-called PCA pyramids. Then we present an image compression method based on the PCA pyramid, which is similar to the Laplace pyramid and wavelet transform. Some experimental results with real images are reported. Finally, we present a method to combine the quantization step with the learning of the PCA pyramid.

7:30 VI:3 UNSUPERVISED CLASSIFICATION OF 3D OBJECTS FROM 2D VIEWS

SATOSHI SUZUKI and HIROSHI ANDO, ATR Human Information Processing Research Laboratories

This paper presents an unsupervised learning scheme for categorizing 3D objects from their 2D projected images. The scheme exploits an auto-associative network's ability to encode each view of a single object into a representation that indicates its view direction. We propose two models that employs different classification mechanisms; the first model selects an auto-associative network whose recovered view best matches the input view, and the second model is based on a modular architecture whose additional network classifies the views by splitting the input space nonlinearly. We demonstrate the effectiveness of the proposed classification models through simulations using 3D wire-frame objects.

7:30 VI:4 FAST ALGORITHMS FOR 2D AND 3D POINT MATCHING: POSE ESTIMATION AND CORRESPONDENCE

STEVEN GOLD, CHIEN PING LU, ANAND RANGARAJAN, SUGUNA PAPPU, and ERIC MJOLSNESS, Yale University

A fundamental open problem in computer vision-determining pose and correspondence between two sets of points in space - is solved with a novel, fast, robust and easily implementable algorithm. The technique works on noisy point sets that may be of unequal sizes and may differ by non-rigid transformations. A 2D variation calculates the pose between point sets related by an affine transformation - translation, rotation, scale and shear. A 3D to 3D variation calculates translation and rotation. An objective describing the problem is derived from mean field theory. The objective is minimized with clocked (EM-like) dynamics. Experiments with both handwritten and synthetic data provide empirical evidence for the method.

WEDNESDAY AM
ORAL SESSION 5

APPLICATIONS

8:30 **O5.1 HANDWRITING RECOGNITION FOR THE NEWTON (INVITED TALK)**
L. KITAINIK, ParaGraph International

9:00 **O5.2 TRANSFORMATION INVARIANT AUTOASSOCIATION WITH APPLICATION TO HANDWRITTEN CHARACTER RECOGNITION**
HOLGER SCHWENK and MAURICE MILGRAM, Universite Pierre et Marie Curie

When training neural networks by the classical backpropagation algorithm the whole problem to learn must be expressed by a set of inputs and desired outputs. However, we often have high-level knowledge about the learning problem. In optical character recognition (OCR), for instance, we know that the classification should be invariant under a set of transformations like rotation or translation. We propose a new modular classification system based on several autoassociative multilayer perceptrons which allows the efficient incorporation of such knowledge. Results are reported on the NIST database of upper case handwritten letters and compared to other approaches to the invariance problem.

9:20 **SPOTLIGHT V: APPLICATIONS**
RECOGNIZING HANDWRITTEN DIGITS USING MIXTURES OF LINEAR MODELS, Geoffrey E. Hinton, Michael Revow, and Peter Dayan, University of Toronto

FACTORIAL LEARNING AND THE EM ALGORITHM, Zoubin Ghahramani, MIT

LEARNING MANY RELATED TASKS AT THE SAME TIME WITH BACKPROPAGATION, Rich Caruana, Carnegie Mellon University

9:30 **O5.3 LEARNING PROTOTYPE MODELS FOR TANGENT DISTANCE**
TREVOR HASTIE, PATRICE SIMARD, and EDUARD SACKINGER, AT&T Bell Laboratories

Simard, LeCun & Denker (1993) showed that the performance of nearest-neighbor classification schemes for handwritten character recognition can be improved by incorporating invariance to specific transformations in the underlying distance metric - the so called *tangent distance*. The resulting classifier, however, can be prohibitively slow and memory intensive due to the large amount of prototypes that need to be stored and used in the distance comparisons. In this paper we develop rich models for representing large subsets of the prototypes. These models are either used singly per class, or as basic building blocks in conjunction with the K-means clustering algorithm.

9:50 **O5.4 REAL-TIME CONTROL OF A TOKAMAK PLASMA USING NEURAL NETWORKS**

CHRIS M. BISHOP, Aston University and PAUL S. HAYNES, MIKE E.U. SMITH, TOM N. TODD and DAVID L. TROTMAN, AEA Technology

This paper presents results from the first use of neural networks for the real-time feedback control of high temperature plasmas in a tokamak fusion experiment. The tokamak is currently the principal experimental device for research into the magnetic confinement approach to controlled fusion. In the tokamak, hydrogen plasmas, at temperatures of up to 100 Million K, are confined by strong magnetic fields. Accurate control of the position and shape of the plasma boundary requires real-time feedback control of the magnetic field structure on a time-scale of a few tens of microseconds. Software simulations have demonstrated that a neural network approach can give significantly better performance than the linear technique currently used on most tokamak experiments. The practical application of the neural network requires high-speed hardware, and this has been achieved by developing a novel, fully parallel implementation of the multilayer perceptron using a hybrid of digital and analog technology.

10:10 **BREAK**

ORAL SESSION 6
IMPLEMENTATION

10:40 **O6.1 ICEG MORPHOLOGY CLASSIFICATION USING AN ANALOGUE VLSI NEURAL NETWORK**

RICHARD COGGINS, MARWAN JABRI, BARRY FLOWER, and STEPHEN PICKARD, University of Sydney

An analogue VLSI neural network has been designed and tested to perform cardiac morphology classification tasks. Analogue techniques were chosen to meet the strict power and area requirements of an Implantable Cardioverter Defibrillator (ICD) system. The robustness of the neural network architecture reduces the impact of noise, drift and offsets inherent in analogue approaches. The network is a 10:6:3 multi-layer perceptron with on chip digital weight storage, a bucket brigade input to feed the Intracardiac Electrogram (ICEG) to the network and has a winner take all circuit at the output. The network was trained in loop and included a commercial ICD in the signal processing path. The system has successfully distinguished arrhythmia for different patients with better than 90% true positive and true negative detections for dangerous rhythms which cannot be detected by present ICDs. The chip was implemented in 1.2um CMOS and consumes less than 200nW maximum average power in an area of $2.2 \times 2.2mm^2$.

11:00 **O6.2 A SILICON AXON**

BRADLEY A. MINCH, PAUL HASLER, CHRIS DIORIO,and CARVER MEAD, California Institute of Technology

We present a silicon model of an axon which shows promise as a building block for pulse-based neural computations involving correlations of pulses across both space and time. The circuit shares a number of features with its biological counterpart including an excitation threshold, a brief refractory period after pulse completion, pulse amplitude restoration, and pulse width restoration. We provide a simple explanation of circuit operation and present data from a chip fabricated in a standard $2\mu m$ CMOS process through the MOS Implementation Service (MOSIS). We emphasize the necessity of the restoration of the width of the pulse in time for stable propagation in axons.

11:20 **SPOTLIGHT VI: IMPLEMENTATIONS**

PREDICTING THE RISK OF COMPLICATIONS IN CORONARY ARTERY BYPASS OPERATIONS USING NEURAL NETWORKS, Richard P. Lippmann and Yuchun Lee, MIT Lincoln Laboratory and Dr. David Shahian, Lahey Clinic

LOCAL ERROR BARS FOR NONLINEAR REGRESSION AND TIME SERIES PREDICTION, David A. Nix and Andreas S. Weigend, University of Colorado

DYNAMIC CELL STRUCTURES, Jorg Bruske and Gerald Sommer, Christian Albrechts University at Kiel, Germany

11:30 **O6.3 THE NI1000: HIGH SPEED PARALLEL VLSI FOR IMPLEMENTING MULTILAYER PERCEPTRONS**

MICHAEL P. PERRONE, Thomas J. Watson Research Center and LEON N. COOPER, Brown University

We present a new version of the standard multilayer perceptron (MLP) algorithm for the state-of-the-art in neural network VLSI implementations: the Intel Ni1000. This approach enables the standard MLP to utilize the parallel architecture of the Ni1000 to achieve on the order of 40000, 256-dimensional classifications per second. Due to the compact size and affordable price of the Ni1000, this classification speed could be available for the average personal computer.

11:50 **O6.4 ANALOG VLSI IMPLEMENTATION OF THE ART1 ALGORITHM**

T. SERRANO, B. LINARES-BARRANCO, and J.L. HUERTAS, National Microelectronics Center, Spain

We describe an analog VLSI implementation of the ART1 algorithm (Carpenter, 1987). A prototype chip has been fabricated in a standard low cost $1.5\mu m$ double-metal single-poly CMOS process. It has a die area of $1cm^2$ and is mounted in a 120-pins PGA package. The chip realizes a modified version of the original ART1 architecture (Serrano, 1994a). Such modification has been shown to preserve all computational properties of the original algorithm (Serrano, 1994b), while being more appropriate for VLSI realizations. The chip implements an ART1 network with 100 $F1$ nodes and 18 $F2$ nodes. It can therefore cluster 100 binary pixels input patterns into up to 18 different categories. Modular expandability of the system is possible by assembling an $N \times M$ array of chips without any extra interfacing circuitry, resulting in an $F1$ layer with $100 \times N$ nodes, and an $F2$ layer with $18 \times M$ nodes. Pattern classification is performed in less than $1\mu s$, which means an equivalent computing power of 1.8×10^9 connections per second. Although internally the chip is analog in nature, it interfaces to the outside world through digital signals, thus having a true asynchrounous digital behavior. Experimental chip test results are available, which have been obtained through test equipments for digital chips.

12:10 **LUNCH**

WEDNESDAY PM
ORAL SESSION 7

SPEECH AND SIGNAL PROCESSING

2:00 **O7.1 CORRELOGRAMS : A TOOL FOR SOUND SEPARATION (INVITED TALK)**
M. SLANEY, Apple Computer

2:30 **O7.2 NON-LINEAR PREDICTION OF ACOUSTIC VECTORS USING HIERARCHICAL MIXTURES OF EXPERTS**
S.R. WATERHOUSE and A.J. ROBINSON, Cambridge University

In this paper we propose the use of the Hierarchical Mixture of Experts (HME) architecture of (Jordan & Jacobs 1994) to perform non-linear prediction of speech within the framework of Vector Predictive Coding (Cuperman & Gersho 1985). By combining vector non-linear prediction with probabilistic information derived from the HME, it is anticipated that low-bit rate speech coding will be achieved. Prediction of speech from the Resource Management (RM) corpus shows superior performance of the HME over linear predictors. We also show that the linear predictors of the HME learn to specialise on different classes of speech vectors.

2:50 **SPOTLIGHT VII: SPEECH AND SIGNAL PROCESSING**

A CONNECTIONIST TECHNIQUE FOR ACCELERATED TEXTUAL INPUT: LETTING A NETWORK DO THE TYPING, Dean A. Pomerleau, Carnegie Mellon University

PREDICTIVE CODING WITH NEURAL NETS: APPLICATION TO TEXT COMPRESSION, Stefan Heil and Jurgen Schmidhuber, Technische Universitat Munchen

HIERARCHICAL MIXTURES OF EXPERTS APPLIED TO A FRAME-BASED NEURAL NETWORK SYSTEM FOR CONTINUOUS SPEECH RECOGNITION, Ying Zhao, Richard Schwartz and John Makhoul, BBN System and Technologies

3:00 **O7.3 GLOVE-TALK II: MAPPING HAND GESTURES TO SPEECH USING NEURAL NETWORKS**

S. SIDNEY FELS and GEOFFREY E. HINTON, University of Toronto

Glove-TalkII is a system which translates hand gestures to speech through an adaptive interface. Hand gestures are mapped continuously to 10 control parameters of a parallel formant speech synthesizer. The mapping allows the hand to act as an artificial vocal tract that produces speech in real time. This gives the unlimited vocabulary in addition to direct control of fundamental frequency and volume. Currently, the best version of Glove-TalkII uses several input devices (including a Cyberglove, a 3-space tracker, a keyboard and a foot-pedal), a parallel formant speech synthesizer and 3 neural networks. The gesture-to-speech task is divided into vowel and consonant production by using a gating network to weight the outputs of a vowel and a consonant neural network. The gating network and the consonant network are trained with examples from the user. The vowel network implements a fixed, user-defined relationship between hand-position and vowel sound and does not require any training examples from the user. Volume, fundamental frequency and stop consonants are produced with a fixed mapping from the input devices. One subject has trained to speak intelligibly with Glove-TalkII. He speaks slowly with speech quality similar to a text-to-speech synthesizer but with far more natural-sounding pitch variations.

3:20 **O7.4 VISUAL SPEECH RECOGNITION WITH STOCHASTIC NETWORKS**

JAVIER R. MOVELLAN, University of California San Diego.

This paper presents developments on a stochastic network for speaker independent visual speech recognition. The approach in this investigation was to feed stochastic networks relatively unprocessed raw images and train them with the EM algorithm. The images were modeled as mixtures of independent radial basis functions, and the temporal dependencies were captured with a standard left-right Markov Process. The system achieved human-like performance when recognizing the first four English digits.

3:40 BREAK

ORAL SESSION 8

VISION

4:15 **O8.1 LEARNING SACCADIC EYE MOVEMENTS USING MULTISCALE SPATIAL FILTERS**

RAJESH P.N. RAO and DANA H. BALLARD, University of Rochester

We describe a framework for learning saccadic eye movements using a photometric representation of target points in natural scenes. The representation takes the form of a long vector comprised of the responses of spatial filters at different orientations and scales. We first demonstrate the use of this response vector in the task of locating previously foveated points in a scene and subsequently use this property in a multisaccade strategy to derive an adaptive motor map for delivering accurate saccades.

4:35 **SPOTLIGHT VIII: VISION**

LEARNING DIRECTION IN GLOBAL MOTION: TWO CLASSES OF PSYCHOPHYSICALLY-MOTIVATED MODELS, V. Sundareswaran and Lucia M. Vaina, Boston University

DECORRELATION DYNAMICS: A THEORY FOR ORIENTATION CONTRAST AND ADAPTATION, Dawei W. Dong, University of California, Berkeley

LIMITS ON LEARNING MACHINE ACCURACY IMPOSED BY DATA QUALITY, Corinna Cortes, L.D. Jackel, and Wan-Ping Chiang, AT&T Bell Laboratories

4:45 **O8.2 A CONVOLUTIONAL NEURAL NETWORK HAND TRACKER**

STEVEN J. NOWLAN and JOHN C. PLATT, Synaptics, Inc.

We describe a system which can track a hand in a sequence of video frames and recognize hand gestures in a user independent manner. The system locates the hand in each video frame and determines if the hand is open or closed. The tracking system is able to track the hand to within ±10 pixels of its correct location in 99.7% of the frames from a test set containing video sequences from 18 different individuals captured in 18 different room environments. The gesture recognition network correctly determines if the hand being tracked is open or closed in 99.1% of the frames in this test set. The system has been designed to operate in real time with existing hardware.

5:05 **O8.4 CORRELATION AND INTERPOLATION NETWORKS FOR REAL-TIME EXPRESSION ANALYSIS/SYNTHESIS**

TREVOR DARRELL, IRFAN ESSA, and ALEX PENTLAND, MIT Media Lab

We describe a framework for real-time tracking of facial expressions that uses neurally-inspired correlation and interpolation methods. A distributed view-based representation is used to characterize facial state, and is computed using a replicated correlation network. The ensemble response of the set of view correlation scores is input to a network based interpolation method, which maps perceptual state to motor control states for a simulated 3-D face model. Activation levels of the motor state correspond to muscle activations in an anatomically derived model. By integrating fast and robust 2-D processing with 3-D models, we obtain a system that is able to quickly track and interpret complex facial motions in real-time.

5:25 **DINNER**

WEDNESDAY EVENING POSTERS

ALGORITHMS & ARCHITECTURES

7:30 AA:21 FACTORIAL LEARNING AND THE EM ALGORITHM

ZOUBIN GHAHRAMANI, MIT

Data is often generated from an interaction of multiple causes or factors. In this paper, we present an unsupervised learning algorithm for extracting the multiple causal structure of such data sets. The algorithm is derived from the maximum likelihood framework of the Expectation-Maximization (EM) procedure. We show that, as a result of the combinatorial nature of the data generation process, the E-step of the standard EM algorithm is computationally intractable. We therefore propose two alternate methods of computing the E-step and relate them to the Boltzmann learning algorithm. Finally, we illustrate the algorithms with a simple simulation and discuss an extension to learning in factorial Markov chains.

7:30 AA:22 A GROWING NEURAL GAS NETWORK LEARNS TOPOLOGIES

BERND FRITZKE, Ruhr-Universitat Bochum

An incremental network model is introduced which is able to learn the important topological relations in a given set of input vectors by means of a simple Hebb-like learning rule. In contrast to previous approaches, our model has no parameters which change over time and it is able to continue learning and adding units and connections until a performance criterion is met. Applications of the model include vector quantization, clustering, interpolation.

7:30 AA:23 LOCAL ERROR BARS FOR NONLINEAR REGRESSION AND TIME SERIES PREDICTION

DAVID A. NIX and ANDREAS S. WEIGEND, University of Colorado

We present a new method for obtaining "local error bars", i.e., estimates of the error of the network outputs that depend on the input. This maximum-likelihood nonlinear-regression technique is first demonstrated on an artificial example with locally-varying normally distributed target noise and then applied to the laser data from the *Santa Fe Time Series Prediction and Analysis Competition*. We then show an extension that allows the estimation of error bars for iterated predictions and apply it to the exact competition task. This pricipled method gives the best performance on the competition task to date.

7:30 AA:24 AN ALTERNATIVE MODEL FOR MIXTURES OF EXPERTS

LEI XU, The Chinese University of Hong Kong, MICHAEL I. JORDAN, MIT, and GEOFFREY E. HINTON, University of Toronto

An alternative model is proposed for mixtures of experts, by utilizing a different parametric form for the gating network. The modified model is trained by an *EM algorithm*. In comparison with earlier models - trained by either EM or gradient ascent - there is no need to select a learning stepsize to guarantee the convergence of the learning procedure. We report simulation experiments which show that the new architecture yields significantly faster convergence. We also apply the new model to two problems domains: piecewise nonlinear function approximation and the combination of multiple previously trained classifiers.

7:30 **AA:25 ESTIMATING CONDITIONAL PROBABILITY DENSITIES FOR PERIODIC VARIABLES**

CHRIS M. BISHOP and CLAIRE LEGLEYE, Aston University

Most of the common techniques for estimating conditional probability densities are inappropriate for applications involving periodic variables. In this paper we introduce two novel techniques for tackling such problems, and investigate their performance using synthetic data. We then apply these techniques to the problem of extracting the distribution of wind vector directions from radar scatterometer data gathered by a remote-sensing satellite.

7:30 **AA:26 EFFECTS OF NOISE ON CONVERGENCE AND GENERALIZATION IN RECURRENT NETWORKS**

KAM JIM, BILL G. HORNE and C. LEE GILES, NEC Research Institute

We introduce and study methods of inserting *synaptic noise* into dynamically-driven recurrent neural networks and show that applying a controlled amount of noise during training may improve convergence and generalization. In addition, we analyze the effects of each noise parameter (additive vs. multiplicative, cumulative vs. non-cumulative, per time step vs. per string) and predict that best overall performance can be achieved by injecting additive noise at each time step. Extensive simulations on learning the dual parity grammar from temporal strings substantiate these predictions.

7:30 **AA:27 LEARNING MANY RELATED TASKS AT THE SAME TIME WITH BACKPROPAGATION**

RICH CARUANA, Carnegie Mellon University

Hinton proposed that generalization in artificial neural nets should improve if nets learn to represent the domain's underlying regularities. Abu-Mustafa's *hints* work suggests that the *outputs* of a backprop net can be thought of as *inputs* through which domain-specific information can be given to the net. We extend these hypotheses by proposing that a backprop net learning many related tasks at the same time can use these tasks as inductive bias for each other and thus learn better. We identify several multitask backprop mechanisms and provide empirical evidence that multitask backprop can yield better generalization in real domains.

7:30 **AA:28 A RAPID GRAPH-BASED METHOD FOR ARBITRARY TRANSFORMATION INVARIANT PATTERN CLASSIFICATION**

ALESSANDRO SPERDUTI, Universita di Pisa and DAVID G. STORK, Ricoh California Research Center

We present a graph-based method for rapid, accurate search through prototypes for transformation invariant pattern classification. Our method has in theory the same recognition accuracy as other recent methods based on "tangent distance" (Simard et al., 1994), since it uses the same categorization rule. Nevertheless ours is significantly faster during classification because far fewer tangent distances need be computed. Crucial to the success of our system are 1) a novel graph architecture in which transformation constraints and geometric relationships among prototypes are encoded during learning, and 2) an improved graph search criterion, used during classification. These architectural insights are applicable to a wide range of problem domains. Here we demonstrate that on a handwriting recognition task, a basic implementation of our system requires less than half the computation of the leading alternate method.

7:30 AA:29 RECURRENT NETWORKS: SECOND ORDER PROPERTIES AND PRUNING

MORTEN WITH PEDERSEN and LARS KAI HANSEN, Technical University of Denmark

Second order properties of cost functions for recurrent networks are investigated. We analyze a layered fully recurrent architecture, the virtue of this architecture is that it features the conventional feedforward as a special case. A detailed description of recursive computation of the *full Hessian* of the network cost function is provided. We discuss the possibility of invoking simplifying approximations of the Hessian and show how weight decays iron the cost function and thereby greatly assist training. We present tentative pruning results, using Hassibi et al.'s *Optimal Brain Surgeon*, demonstrating that recurrent networks can construct an efficient internal memory.

7:30 AA:30 CLASSIFYING WITH GAUSSIAN MIXTURES, CLUSTERS, AND SUBSPACES

NANDA KAMBHATLA and TODD K. LEEN, Oregon Graduate Institute of Science & Technology

This paper develops the relationship between Bayes classifiers that use Gaussian mixtures to model class-conditional densities, and several non-Bayesian classification algorithms that use a quadratic distance measure as the discriminant function. Examples include cluster-based algorithms such as learning vector quantization, and subspace classifiers based on principal component analysis. The analysis suggests several new algorithms that improve the performance of existing techniques. We show empirical results for a phoneme recognition task.

7:30 AA:31 EFFICIENT METHODS FOR DEALING WITH MISSING DATA IN SUPERVISED LEARNING

VOLKER TRESP, RALPH NEUNEIER and SUBUTAI AHMAD, Siemens AG

We present efficient methods for dealing with the problem of missing inputs during training and recall. For recall, we obtain closed form solutions for arbitrary feedforward networks. A similar solution is found for the error gradient in training. We verify our theoretical results using a classification problem.

7:30 AA:32 AN EXPERIMENTAL COMPARISON OF RECURRENT NEURAL NETWORKS

BILL G. HORNE and C. LEE GILES, NEC Research Institute

Many different discrete-time recurrent neural network architectures have been proposed. However, there has been virtually no effort to compare these architectures experimentally. In this paper we review and categorize many of these architectures and compare how they perform on various classes of simple problems including grammatical inference and nonlinear system identification.

7:30 AA:33 ACTIVE LEARNING WITH STATISTICAL MODELS

DAVID A. COHN, ZOUBIN GHAHRAMANI, and MICHAEL I. JORDAN, MIT

For many types of learners one can compute the statistically "optimal" way to select data. We review how these techniques have been used with feedforward neural networks (MacKay, 1992; Cohn, 1994). We then show how the same principles may be used to select data for two alternative, statistically-based learning architectures: Mixtures of Gaussians and locally weighted regression. While the techniques for neural networks are expensive and approximate, the techniques for mixtures of Gaussians and locally weighted regression are both efficient and accurate.

7:30 AA:34 DYNAMIC CELL STRUCTURES

JORG BRUSKE and GERALD SOMMER, Christian Albrechts University at Kiel

Dynamic Cell Structures (DCS) are a flexible ANN architecture merging and extending the ideas of B. Fritzke and T. Martinetz. DCS can be used both for unsupervised and supervised learning. In the former case, DCS learn to build perfectly topology preserving feature maps by inserting/deleting neural units as well as employing a modified Kohonen learning rule in conjunction with competitive Hebbian learning. The Kohonen type learning rule serves to adjust the synaptic weight vectors while Hebbian learning establishes a lateral connection structure between the units reflecting the topology of the feature manifold. In case of supervised learning, i.e. function approximation, each neural unit implements a Radial Basis Function, and an additional layer of linear output units adjusts according to the delta-rule. Insertion/deletion of units is guided by the local approximation error and the emerging lateral connection structure, leading to a very efficient approximation scheme both in terms of computational effort per example and total number of example presentations. Superiority to conventional methods is demonstrated by applying DCS to a couple of CMU Benchmark tests.

7:30 AA:35 LEARNING WITH PREKNOWLEDGE: CLUSTERING WITH POINT AND GRAPH MATCHING DISTANCE MEASURES

STEVEN GOLD, ANAND RANGARAJAN and ERIC MJOLSNESS, Yale University

Prior constraints are imposed upon a learning problem in the form of distance measures. Prototypical 2-D point sets and graphs are learned by clustering with point matching and graph matching distance measures. The point matching distance measure is invariant under affine transformations - translation, rotation, scale and shear - and permutations. It operates between noisy images with missing and spurious points. The graph matching distance measure operates on weighted graphs and is invariant under permutations. Learning is formulated as an optimization problem. Large objectives so formulated (~ million variables) are efficiently minimized using a combination of optimization techniques - algebraic transformations, projection methods, clocked objectives, and deterministic annealing.

7:30 AA:36 SETTLING TEMPORAL DIFFERENCES: TIME SERIES PREDICTION USING TD (λ)

PETER T. KAZLAS and ANDREAS S. WEIGEND, University of Colorado

We apply the paradigm of Temporal Difference (TD) learning (Sutton, 1988) to forecasting the behavior of dynamical systems with real-valued outputs (as opposed to game-like situations, Tesauro, 1992). Prediction nonlinear dynamical systems requires individual networks with nonlinear hidden units as individual predictors. In this paper, we compare TD learning with supervised learning, both theoretically and on a real-world example, on the laser data from the Santa Fe Competition. For both paradigms, we use two architectures; the first architecture ("separate hidden units") consists of individual networks for each of the five direct multi-step prediction tasks, the second ("shared hidden units") has a single (larger) layer of hidden units that generates a representation that is used to generate all five predictions for the next five steps. We find that standard supervised learning outperforms TD learning in the case of separate hidden units, but TD learning outperforms standard supervised learning when the hidden units are shared.

APPLICATIONS

7:30 AP:21 COMPARING THE PREDICTION ACCURACY OF ARTIFICIAL NEURAL NETWORKS AND OTHER STATISTICAL MODELS FOR BREAST CANCER SURVIVAL

HARRY B. BURKE, DAVID B. ROSEN, and PHILIP H. GOODMAN, University of Nevada School of Medicine

Background. The TNM staging system has been used since the early 1960's to predict breast cancer patient outcome. In an attempt to increase prognostic accuracy, many putative prognostic factors have been identified. Because the TNM stage model can not accommodate these new factors, the proliferation of factors in breast cancer has lead to clinical confusion. What is required is a new computerized prognostic system that can test putative prognostic factors and integrate the predictive factors with the TNM variables in oreder to increase prognostic accuracy.

Methods. Using the area under the curve of the receiver operating characteristic, we compare the accuracy of the following predictive models in terms of five year breast cancer-specific survival:pTNM staging system, principal component analysis, classification and regression trees, logistic regression, cascade correlation neural network, conjugate gradient descent neural network, probabilistic neural network, and backpropagation neural network.

Results. Several statistical models are significantly more accurate than the TNM staging system. Logistic regression and the backpropagation neural network are the most accurate prediction models for predicting five year breast cancer-specific survival.

Conclusions. Computerized prediction systems such as logistic regression and artificial neural networks are more accurate than the current look-up table system. In addition, artificial neural networks have the potential to discover nonmonotonicity and complex interactions without a priori information.

7:30 AP:22 A CONNECTIONIST TECHNIQUE FOR ACCELERATED TEXTUAL INPUT: LETTING A NETWORK DO THE TYPING

DEAN A. POMERLEAU, Carnegie Mellon University

Each year people spend a huge amount of time typing. The text people type typically contains a tremendous amount of redundancy due to predictable word usage patterns and the text's structure. This paper describes a neural network system call AutoTypist that monitors a person's typing and predicts what will be entered next. AutoTypist displays the most likely subsequent word to the typist, who can accept it with a single keystroke, instead of typing it in its entirety. The multi-layer perceptron at the heart of AutoTypist adapts its predictions of likely subsequent text to the user's word usage pattern, and to the characteristics of the text currently being typed. Increases in typing speed of 4-8% when typing English prose and 10-20% when typing C code have been demonstrated using the system, suggesting a potential time savings of more than 20 hours per user per year.

7:30 **AP:23 LEARNING TO PLAY THE GAME OF CHESS**
SEBASTIAN THRUN, University of Bonn

This paper presents initial results with NeuroChess, a tool for learning to play chess from the final outcome of games. NeuroChess learns chess board evaluation functions, represented by artificial neural networks. The central learning mechanism of the NeuroChess approach integrates inductive neural network learning, temporal differencing, and a variant of explanation-based learning. Thus far, NeuroChess has managed to defeat GNU-Chess, a publicly available chess tool, in hundreds of games.

7:30 **AP:24 PREDICTIVE CODING WITH NEURAL NETS: APPLICATION TO TEXT COMPRESSION**
STEFAN HEIL and JURGEN SCHMIDHUBER, Technische Universitat Munchen

We demonstrate that neural networks are promising tools for discrete data compression without loss of information. We combine predictive neural nets and standard statistical predictive coding techniques to compress text files. Tested on short German newspaper articles, our method clearly outperforms the widely used Lempel-Ziv algorithm (which is asymptotically optimal and builds the basis of the UNIX functions "compress" and "gzip").

7:30 **AP:25 PREDICTING THE RISK OF COMPLICATIONS IN CORONARY ARTERY BYPASS OPERATIONS USING NEURAL NETWORKS**
RICHARD P. LIPPMANN and YUCHUN LEE, MIT Lincoln Laboratory and DR. DAVID SHAHIAN, Lahey Clinic

Accurate estimates of the risks involved in medical procedures can be used to compare quality of care across institutions, to provide advice to individual patients, and to gain insight into patient or procedural characteristics that have the greatest impact on success. Initial experiments have demonstrated that sigmoid multilayer perceptron networks provide better risk prediction than more conventional logistic regression when used to predict the risk of death and stroke on 791 patients who underwent coronary artery bypass operations at the Lahey Clinic. A multilayer sigmoid network provided significantly better risk prediction across all subjects than logistic regression when both algorithms used the same input features and training and testing data. This network provided a sensitivity (detection rate for fatal complications) of roughly 65% when the specificity (100 - false alarm rate for normal patients) was roughly 80%. All testing was performed using 10-fold cross validation. These encouraging results are currently being validated using a larger data base and approaches to determining the confidence of risk prediction for individual patients are being explored.

7:30 **AP:26 A MIXTURE MODEL NEURAL EXPERT SYSTEM FOR DIAGNOSIS**
MAGNUS STENSMO and TERRENCE J. SEJNOWSKI, Salk Institute

A framework where diagnosis is viewed as classification with missing data is presented. The data is modeled by a mixture of Gaussians whose parameters are estimated by the EM algorithm. Regression gives a current diagnosis in view of observed data and finds a suitable question to ask to obtain new information. This is repeated until a conclusive result is reached. A system with this functionality has been built and results when applied to a heart disease database are presented. The system can handle missing data both when training and classifying, where a feed-forward multi-layer perceptron would have problems. It is also domain independent and the time needed for system construction is very low compared to traditional expert systems since no knowledge engineering is needed.

7:30 AP:27 INFERRING GROUND TRUTH FROM SUBJECTIVE LABELLING OF VENUS RADAR IMAGES

P. SMYTH, M. BURL, U.M. FAYYAD, P. BALDI, Jet PropulsionLaboratory and P. PERONA, California Institute of Technology

In remote sensing applications "ground-truth" data is often used as the basis for estimating spatial statistics or for training pattern recognition algorithms to generate thematic maps or detect objects of interest. In reality, getting verifiable accurate ground truth is often either prohibitively expensive or physically impossible. Instead one must often rely on the subjective opinions of experts who can visually examine the images and provide a subjective labelling, in essence, noisy estimates of the "true" ground truth. Of particular importance is the ability to calibrate the reliability and bias of individual labellers: this is a non-trivial problem. In addition, the problem of combining multiple opinions is also important. In this paper we discuss some of our recent work on this topic in the context of detecting small volcanoes in Magellan SAR images of Venus. Experimental results (using the Expectation-Maximization algorithm) suggest that accounting for subjective noise can be surprisingly important in terms of quantifying both human and algorithm detection performance.

CHARACTER RECOGNITION

7:30 CR:21 THE USE OF DYNAMIC WRITING INFORMATION IN A CONNECTIONIST ON-LINE CURSIVE HANDWRITING RECOGNITION SYSTEM

STEFAN MANKE and MICHAEL FINKE, University of Karlsruhe, and ALEX WAIBEL, Carnegie Mellon University

Writer independent, large vocabulary on-line handwriting recognition systems require robust input representations and recognition techniques making optimal use of dynamic writing information, i.e. the time-ordered coordinate sequence written on a graphics tablet. In this paper we describe an input representation for cursive handwriting, which combines this dynamic writing information with static bitmaps used on optical character recognition, and propose a connectionist recognizer, which integrates segmentation and recognition in a single framework. This connectionist recognizer, a so called Multi-State Time Delay Neural Network (MS-TDNN), is well suited for handling temporal sequences of patterns as provided by this kind of input representation. Our system has been tested both on different single character recognition tasks and large vocabulary, cursive handwriting recognition tasks with vocabulary sizes up to 20000 words. We achieved recognition rates up to 99.5% on writer independent, single character recognition tasks and up to 98.1% on writer dependent, cursive handwriting tasks.

7:30 CR:22 ADAPTIVE ELASTIC INPUT FIELD FOR RECOGNITION IMPROVEMENT

MINORU ASOGAWA, C&C Systems Research Laboratories, NEC

For machines to perform classification tasks, such as speech and character recognition, appropriately handling deformed patterns is a key to achieving high performance. We present a new type of classification system, an Adaptive Input Field Neural Network (AIFNN), which includes a simple pre-trained neural network and an elastic input field attached to an input layer. By using an iterative method, AIFNN can determine an optimal AIFNN translation for an elastic input field to compensate the original deformations. The convergence of the AIFNN algorithm is shown. AIFNN is applied for a handwritten numerals recognition. Consequently 10.83% of originally misclassified patterns are correctly categorized and total performance is improved without modifying the neural network.

7:30 CR:23 RECOGNIZING HANDWRITTEN DIGITS USING MIXTURES OF LINEAR MODELS

GEOFFREY E. HINTON, MICHAEL REVOW, and PETER DAYAN, University of Toronto

We construct a set of logically linear generative models of a collection of pixel-based images of digits, and use them for recognition. Different models of a given digit are used to capture different styles of writing, and new images are classified by evaluating their log-likelihoods under each model. We use an EM-based algorithm in which the M-step is computationally straightforward principal components analysis (PCA). Incorporating tangent-prop information (Simard *etal* 1992) about expected local deformations only requires adding tangent vectors into the sample covariance matrices for the PCA, and it demonstrably iimproves performance.

7:30 CR:24 PAIRWISE NEURAL NETWORK CLASSIFIERS WITH PROBABILISTIC OUTPUTS

DAVID PRICE, STEFAN KNERR,LEON PERSONNAZ, and GERARD DREYFUS, ESPCI

Multi-class classification problems can be efficiently solved by partitioning the original problem into sub-problems involving only two classes: for each pair of classes, a (potentially small) neural network is trained using only the data of these two classes. The outputs of all the two-class networks can be combined to give binary decisions for the class labels by use of simple logic operations. However, many pattern recognition applications ask for probabilistic class decisions which may be used subsequently in higher context levels. A prominent example is speech or handwriting recognition, where the probabilities at the output of the phoneme or character recognizer are often used by a Hidden-Markov-Model or some other dynamic programming algorithm to compute probabilities for word hypothesis. In this paper, we show how to combine the outputs of the two-class neural networks in order to obtain posterior probabilities for the class decisions. The resulting probabilistic pairwise classifier is part of a handwriting recognition system which is currently applied to check reading. We present results on real world data bases and show that, from a practical point of view, these results compare favorably to other neural network approaches.

CONTROL

7:30 CN:21 FORMATION OF INTERNAL MODELS FOR LEARNING CONTROL OF ARM MOVEMENTS

REZA SHADMEHR, TOM BRASHERS-KRUG, and FERDINANDO MUSSA-IVALDI, MIT.

We consider the problem of how the CNS learns to control dynamics of a mechanical system. We show that as humans learn to control their arm movements in a novel dynamical environment, their motor control system changes by an amount which approximates a map from states of the arm to forces imposed by the environment. This map, called an internal model, is presumably implemented via a population of neurons and learning has resulted in changes in the synaptic strengths of the cells in the population. We hypothesize that after learning dynamics of a given environment, the position of the weights has biased this population and should dictate the learning rate for a second environment. We find that subjects have great difficulty learning a second environment if they had just learned an uncorrelated environment. We show that these results may be explained in terms of distances of the two environments in the weight space of a population of neurons.

7:30 CN:22 COMPUTATIONAL STRUCTURE OF COORDINATE TRANSFORMATIONS: A GENERALIZATION STUDY

ZOUBIN GHAHRAMANI, DANIEL M. WOLPERT, and MICHAEL I. JORDAN, MIT

One of the fundamental properties that both neural networks and the central nervous system (CNS) share is the ability to learn and generalize from examples. While this property has been studied extensively in the neural network literature it has not been fully explored in human learning. We have chosen a coordinate transformation system - the inverse kinematic map that transforms visual coordinates into motor coordinates - to study the generalization effects of learning new input-output pairs. In this system, using a paradigm of computer controlled altered visual feedback, we are able to restrict learning to single input-output pairs and can thereby examine subsequent generalization. The results of exposure to either one or two new input-output pairs suggest that the kinematic map generalizes linearly in Cartesian space. The extent of generalization indicates that the transformation is represented globally with the greatest change seen at the training points and an effect that decreases with distance from the remapped input. This study provides constraints on the type of network which could represent such a mapping.

COGNITIVE SCIENCE

7:30 **CS:21 A SOLVABLE CONNECTIONIST MODEL OF IMMEDIATE RECALL OF ORDERED LISTS**

NEIL BURGESS, UCL, London

A model of short-term memory for serially ordered lists is proposed. The network's main characteristics are a decaying Hebbian association from a locally-overlapping representation of an item's context to a representation of its phonemic composition, plus sequential selection and suppression of each item. An approximate mathematical analysis of error probabilities in the presence of Gaussian noise is presented. The model provides a parsimonious explanation for the probability of error in immediate recall as a function of serial position, list length, word length, phonemic similarity, and list familiarity. Extension to a model of the 'articulatory loop' and the related clinical and developmental data is discussed.

IMPLEMENTATIONS

7:30 **IM:21 AN ANALOG NEURAL NETWORK INSPIRED BY FRACTAL BLOCK CODING**

FERNANDO J. PINEDA and ANDREAS G. ANDREOU, The Johns Hopkins University

The fractal block coding approach to compression is summarized and a subthreshold current-mode MOS circuit motivated by this approach is presented. The resulting system is linear at steady state, but has nonlinear dynamics. A novel aspect of the neural network is that the biases of individual neurons depend on their position in the network. Given a set of input parameters (encoded data), the network relaxes to a steady state where the currents coming out of the neurons represent the decoded vector. Essentially, the network solves $\boldsymbol{I} = \boldsymbol{WI} + \boldsymbol{B}$ for $\boldsymbol{I}$, given a sparse parameterized matrix weigh matrix $\boldsymbol{W}$ and a dense parameterized bias vector $\boldsymbol{B}$. We present preliminary experimental data from a test chip implemented in $2\mu m$ CMOS. This chip generates curves with qualitatively fractal shapes.

7:30 **IM:22 A STUDY OF PARALLEL PERTURBATIVE GRADIENT DESCENT**

D. LIPPE and J. ALSPECTOR, Bellcore

We have continued our study of a parallel perturbative learning method (Alspector et al., 1993) and implications for its implementation in analog VLSI. Our new results indicate that, in most cases, a single parallel perturbation of the function parameters (weights in a neural network) is theoretically the best course. This is not true, however, for certain problems and may not generally be true when faced with issues of implementation such as limited precision. In these cases, multiple parallel perturbations may be best as indicated in our previous results.

7:30 IM:23 IMPLEMENTATION OF NEURAL HARDWARE WITH THE NEURAL VLSI OF URAN IN APPLICATIONS OF REDUCED REPRESENTATIONS

IL-SONG HAN and YOUNG-JAE CHOI, Korea Telecom Research Center and KI-CHUL KIM and HWANG-SOO LEE, Korea Advanced Institute of Science and Technology

This paper describes a way of neural hardware implementation with the analog-digital mixed neural chip. The full custom neural VLSI of Universally Reconstructable Artificial Neural-network (URAN) is used to implement speech recognition system and real time control electronics. A multi-layer perceptron with piecewise linear hidden and output neurons are trained with limited accuracy computation with success. The network including a large frame input layer with URAN are used to recognize a digital syllable at a forward retrieval. It is also evaluated for the use in servo control to yield the 70% of improvement with URAN. It is suggested for multichip hardware module (with eight chips or more) to extend the performance and capacity.

7:30 IM:24 SINGLE TRANSISTOR LEARNING SYNAPSES

PAUL HASLER, CHRIS DIORIO, BRADLEY A. MINCH and CARVER MEAD, California Institute of Technology

We describe single-transistor silicon synapses that compute, learn, and provide non-volatile memory retention. The synapses efficiently use the physics of silicon to perform local computations. Learning is either two- or four- quadrant depending upon the circuit operation of the device. The small size of single transistor synapses allows the development of dense synaptic arrays. Memory is accomplished via charge storage on polysilicon floating gates, providing long-term retention without refresh. The synapses operate in the low power subthreshold regime, with weight increases using tunneling and weight decreases using hot electron injection. We present two different implementations of single transistor synapses, and discuss some of the tradeoffs between them. Both devices have been fabricated in the standard $2\mu m$ double - poly, analog process available from MOSIS.

LEARNING THEORY

7:30 LT:21 LIMITS ON LEARNING MACHINE ACCURACY IMPOSED BY DATA QUALITY

CORINNA CORTES, L.D. JACKEL and WAN-PING CHIANG, AT&T Bell Laboratories

Random errors and insufficiencies in databases limit the performance of any classifier trained from and applied to the database. In this paper we propose a method to estimate the limiting performance of classifiers imposed by the database. We demonstrate this technique on the task of predicting failure in telecommunication paths.

7:30 **LT:22 LEARNING FROM QUERIES FOR MAXIMUM INFORMATION GAIN IN UNLEARNABLE PROBLEMS**

PETER SOLLICH and DAVID SAAD, University of Edinburgh

We study the problem of learning to approximate a binary perceptron with a linear perceptron, using training examples generated by queries which maximize the information gain (i.e., minimize the entropy) in student or teacher space. Comparing the results to training on random examples, we find that minimum student space entropy queries lead to the same relative improvement in generalization performance over random examples that would be obtained if the teacher was a noisy linear perceptron. Minimum teacher space entropy queries, on the other hand, lead to a higher generalization error than random examples. These results provide some justification for a Bayesian approach to query learning for maximum information gain.

7:30 **LT:23 BIAS, VARIANCE AND THE COMBINATION OF LEAST SQUARES ESTIMATORS**

RONNY MEIR, Technion

We consider the effect of combining several least squares estimators on the expected performance of a regression problem. Computing the exact bias and variance curves as a function of the sample size we are able to quantitatively compare the effect of the combination on the bias and variance separately, and thus on the expected error which is the sum of the two. Our exact calculations, demonstrate that the combination of estimators is particularly useful in the case where the data set is small and noisy and the function to be learned is unrealizable. For large data sets the single estimator produces superior results. Finally, we show that by splitting the data set into several independent parts and training each estimator on a different subset, the performance can in some cases be significantly improved.

7:30 **LT:24 ON-LINE LEARNING OF DICHOTOMIES**

N. BARKAI and H. SOMPOLINSKY, The Hebrew University and H.S. SEUNG, AT&T Bell Laboratories

The performance of on-line algorithms for learning dichotomies is studied. In on-line learning, the number of examples P is equivalent to the learning time, since each example is presented only once. The learning curve, or generalization error as a function of P, depends on the schedule at which the learning rate is lowered. For a target that is a perceptron rule, the learning curve of the perceptron algorithm can decrease as fast as P^{-1}, if the schedule is optimized. If the target is not realizable by a perceptron, the perceptron algorithm does not generally converge to the solution with lowest generalization error. For the case of unrealizability due to a simple output noise, we propose a new on-line algorithm for a perceptron yielding a learning curve that can approach the optimal generalization error as fast as $P^{-1/2}$. We then generalize the perceptron algorithm to any class of thresholded smooth functions learning a target from that class. For "well-behaved" input distributions, if this algirithm converges to the optimal solution, its learning curve can decrease as fast as P^{-1}.

7:30 LT:25 DYNAMIC MODELLING OF CHAOTIC TIME SERIES WITH NEURAL NETWORKS

JOSE C. PRINCIPE and JYH-MING KUO, University of Florida, Gainesville

This paper discusses the use of artificial neural networks for dynamic modelling of time series. We briefly present the theoretical basis for the modelling as a prediction of a vector time series in reconstructed space. The issues of implementing and training an ANN based predictor constitutes the bulk of the paper. We argue that multistep prediction is more appropriate to capture the dynamics, because it constrains the iterated model. We show how this method can be implemented by a recurrent ANN trained with trajectory learning. We also show how to select the trajectory length to train the iterated predictor for the case of non-chaotic and chaotic time series. Experimental results corroborate the proposed method.

7:30 LT:26 A RIGOROUS ANALYSIS OF LINSKER'S HEBBIAN LEARNING NETWORK

JIANFENG FENG, Universitat Tubingen, and HONG PAN and VWANI P. ROYCHOWDHURY, Purdue University

We propose a novel approach for a rigorous analysis of the nonlinear asymmetric dynamics of Linsker's unsupervised Hebbian learning network. Our analysis allows us to determine the whole set of fixed point attractors of the synaptic stabilization process, and explicitly obtain a necessary and sufficient condition for the emergence of structured receptive fields. These results provide for the first time comprehensive explanations of the generation of the various structured connection patterns and of the roles of the different system parameters of the model, in particular the crucial role of the synaptic density function. Our theoretical predictions are confirmed by numerical simulations.

730 LT:27 SAMPLE SIZE REQUIREMENTS FOR FEEDFORWARD NEURAL NETWORKS

MICHAEL J. TURMON and TERRENCE L. FINE, Cornell University

We address the question of how many training samples are required to ensure that the performance of a neural network of given complexity on its training data matches that obtained when fresh data is applied to the network. This desirable property may be termed "reliable generalization." Well-known results of Vapnik give conditions on the number of training samples sufficient for reliable generalization, but these are higher by orders of magnitude than practice indicates; other results in the mathematical literature involve unknown constants and are useless for our purposes.

This work seeks to narrow the gap between theory and practice by transforming the problem into one of determining the distribution of the supremum of a Gaussian random field in the space of weight vectors, which in turn is attacked by application of a technique called the Poisson clumping heuristic. The idea is that mismatches between training set error and true error occur not for an isolated network but for a group of similar networks. The size of this group of equivalent networks is obtained, and means of computing the size based on the training data are considered. It is shown that in some cases the Poisson clumping technique yields estimates of sample size having the same functional form as earlier ones, but since the new estimates incorporate specific characteristics of the network architecture and data distribution it is felt that more realistic estimates will result.

7:30 **LT:28 ASYMPTOTICS OF GRADIENT-BASED NEURAL NETWORK TRAINING ALGORITHMS**

SAYANDEV MUKHERJEE and TERRENCE L. FINE, Cornell University

We study the symptotic properties of the sequence of iterates of weight-vector estimates obtained by training a multilayer feedforward neural network with a basic gradient-descent method using a fixed learning constant and no batch-processing. In the one-dimensional case, an exact analysis establishes the existence of a limiting distribution that is not Gaussian in general. For the general case and small learning constant, a linearization approximation permits the application of results from the theory of random matrices to again establish the existence of a limiting distribution. We study the first few moments of this distribution to compare and contrast the results of our analysis with those of techniques of stochastic approximation.

NEUROSCIENCE

7:30 **NS:21 SHORT-TERM ACTIVE MEMORY, INHIBITION, AND NEUROMODULATION: A COMPUTATIONAL MODEL OF PREFRONTAL CORTEX FUNCTION**

TODD S. BRAVER and JONATHAN D. COHEN, Cargegie Mellon University and DAVID SERVAN-SCHREIBER, University of Pittsburgh

Accumulating data from neurophysiology and neuropsychology have suggested two information processing roles for prefrontal cortex: a) short-term active memory and b) inhibition. We present a new behavioral task and a computational model which were developed in parallel. The task was developed to probe both of these prefrontal functions simultaneously, and produces a rich set of behavioral data that act as constraints on the model. The model is implemented in continuous-time, thus providing a natural framework in which to study the temporal dynamics of processing in the task. We show how the model can be used to examine the behavioral consequences of neuromodulation. Specifically, we use the model to make novel and testable predictions regarding the behavioral performance of schizophrenics, who are hypothesized to suffer from reduced neuromodulatory tone in prefrontal cortex.

7:30 NS:22 A NEURAL MODEL OF DELUSIONS AND HALLUCINATIONS IN SCHIZOPHRENIA

EYTAN RUPPIN and JAMES A. REGGIA, University of Maryland and DAVID HORN,Tel Aviv University

We implement and study a computational model of Stevens' (1992) neurobiological theory of the pathogenesis of schizophrenia. This theory hypothesizes that the onset of schizophrenia is associated with degeneration to temporal lobe neurons projecting on frontal areas, followed by frontal synaptic regeneration. The attractor neural network model we study represents a frontal module. We analyze how, in the face of weakened external input projections, compensatory strengthening of internal synaptic connections and increased noise levels can maintain memory capacities (which are generally preserved in schizophrenia). These compensatory changes were found to have adverse side effects, reminiscent of the delusions seen in schizophrenia: spontaneous, stimulus-independent retrieval of stored memories is generated, concentrated on just few of the stored patterns. These findings account for the occurrence of schizophrenic delusions and hallucinations without any apparent external trigger, and for their tendency to concentrate on a few central cognitive and perceptual themes. Our results explain why these symptoms tend to wane as schizophrenia progresses, why delayed therapeutical intervention leads to a much slower response, and why delusions and hallucinations may persist for a long duration when they occur.

7:30 NS:23 SPATIAL REPRESENTATIONS IN THE PARIETAL CORTEX MAY USE BASIS FUNCTIONS

ALEXANDRE POUGET and TERRENCE J. SEJNOWSKI, The Salk Institute

The problem of spatial perception is often thought as a problem of coordinate change, which has led to the view that the parietal cortex represents the egocentric positions of objects. We show here that the responses of single parietal neurons are inconsistent with this hypotheses. We approach the problem from the perspective of sensori-motor transformation. In most cases, issuing a motor command in response to a stimulus requires a nonlinear transformation of the incoming sensory signals and, therefore, involves approximating nonlinear functions. The response of single parietal neurons appears to be particularly well-adapted to this task. Their tuning curves can be modeled as a gaussian of retinal position multiplied by a sigmoid of eye position, which is a basis function. We show here how these basis functions can be used to generate receptive fields in retinotopic or head-centered coordinates by a simple linear transformation. This raises the possibility that the parietal cortex does not attempt to compute the positions of objects in a particular frame of reference but instead computes a general purpose representation of the retinal and eye position from which any transformation can be synthesized by direct projection. This representation predicts that hemineglect, a neurological syndrome produced by parietal lesions, should not be confined to egocentric coordinates, but should be observed in multiple frames of reference in single patients, a prediction supported by several experiments.

7:30 **NS:24 GROUPING COMPONENTS OF THREE-DIMENSIONAL MOVING OBJECTS IN AREA MST OF VISUAL CORTEX**

RICHARD S. ZEMEL and TERRENCE J. SEJNOWSKI, The Salk Institute

Many cells in the dorsal part of the medial superior temporal (MSTd) area of visual cortex respond selectively to combinations of expansion/contraction and rotation motions. It has been suggested that these MST neurons respond to flow fields that convey information about the heading of a moving observer. Models based on this hypothesis have considered cell responses in the limited condition of an observer moving through a static environment. Our model is based on an alternative hypothesis, that MSTd is responsible for segmenting moving objects and that selective tuning of MSTd cells reflects the grouping of object components undergoing coherent motion. Such a grouping operation is essential in interpreting scenes containing multiple objects, each with its own motion based on its three-dimensional (3-D) position and velocity relative to the observer. Inputs to the model were generated from sequences of ray-traced images that contained a variety of shapes undergoing independent 3-D motion under different lighting conditions and settings. The input representation was modeled after response properties of neurons in area MT which provides the primary input to area MST. After applying an unsupervised learning algorithm, the units became tuned to patterns signalling coherent motion. The results match many of the known properties of MSTd cells and are consistent with recent studies indicating that these cells process 3-D object motion information.

7:30 **NS:25 A MODEL OF THE NEURAL BASIS OF THE RAT'S SENSE OF DIRECTION**

WILLIAM E. SKAGGS, JAMES J. KNIERIM, HEMANT S. KUDRIMOTI, and BRUCE L. MCNAUGHTON, University of Arizona, Tucson

In the last decade the outlines of the neural structures subserving the sense of direction have begun to emerge. Several investigations have shed light on the effects of vestibular input and visual input on the head direction representation. In this paper, a model is formulated of the neural mechanisms underlying the head direction system. The model is built out of simple ingredients, depending on nothing more complicated than connectional specificity, attractor dynamics, Hebbian learning, and sigmoidal nonlinearities, but it behaves in a sophisticated way and is consistent with most of the observed properties of real head direction cells. In addition it makes a number of predictions that ought to be testable by reasonably straightforward experiments.

SPEECH RECOGNITION

7:30 SP:21 HIERARCHICAL MIXTURES OF EXPERTS APPLIED TO A FRAME-BASED NEURAL NETWORK SYSTEM FOR CONTINOUS SPEECH RECOGNITION

YING ZHAO, RICHARD SCHWARTZ, and JOHN MAKHOUL, BBN System and Technologies

For the past few years, we developed the concept of the Segmental Neural Net (SNN) and a paradigm of combining the SNN with the conventional Hidden Markov Models (HMM) for continuous speech recognition. Recently, we switched to a new paradigm of integrating neural nets into the HMM system. In this new paradigm, a frame-based neural net density estimation is used directly in the HMM system, while in the old paradigm, a sigment-based neural net system is combined with HMM only at the N-best rescoring level. Within the structure of this new paradigm, we implemented a more complicated neural net unit based on the idea of *hierarchical mixtures of experts* in order to improve the neural net training. The method of hierarchical mixtures of experts is a generalization of decision trees for classification and regression using "soft" decision boundaries which can be adjusted by a maximum likelihood technique, the Expectation-Maximization (EM) algorithm. In this paper, we will first give an overview of our frame-based neural net system. Then we will address our version of hierarchical mixtures of experts specifically. We will report some initial results on testing the new system on 5,000-word Wall Street Journal (WSJ) corpus.

VISION

7:30 VI:21 LEARNING DIRECTION IN GLOBAL MOTION: TWO CLASSES OF PSYCHOPHYSICALLY-MOTIVATED MODELS

V. SUNDARESWARAN and LUCIA M. VAINA, Boston University

Perceptual learning is defined as fast improvement in performance and retention of the learned ability over a period of time. In a set of psychophysical experiments we demonstrated that perceptual learning occurs for the discrimination of direction in stochastic motion stimuli. Here we model this learning using two approaches: a clustering model that learns to *accommodate* the motion noise, and an averaging model that learns to *ignore* the noise. We present simulation results showing that the models' performance is consistent with the psychophysical results.

7:30 VI:22 USING A NEURAL NET TO INSTANTIATE A DEFORMABLE MODEL

CHRISTOPHER K.I. WILLIAMS, MICHAEL D. REVOW, and GEOFFREY E. HINTON, University of Toronto

Deformable models are an attractive approach to recognizing non-rigid objects which have considerable within class variability. However, there are severe search problems associated with fitting the models to data. We show that by using neural networks to provide better starting points, the search time can be significantly reduced. The method is demonstrated on a character recognition task.

7:30 **VI:23 DECORRELATION DYNAMICS: A THEORY FOR ORIENTATION CONTRAST AND ADAPTATION**

DAWEI W. DONG, University of California, Berkeley

We examine the implication of the hypothesis that the intracortical connections dynamically decorrelate activities of orientation selective cells. We show that this decorrelation dynamics leads to quantitative predictions of orientation contrast and orientation adaptation which are in good agreement with various experiments.

7:30 **VI:24 NONLINEAR IMAGE INTERPOLATION USING SURFACE LEARNING**

CHRISTOPH BREGLER, University of California, Berkeley and STEPHEN M. OMOHUNDRO, Int. Computer Science Institute

We present the problem of interpolating between specified images in an image sequence as a simple, but important task in model-based vision. We study an approach based on the abstract task of "surface learning" and present results on both synthetic and real image sequences. This is of specific interest for our combined lip-reading and speech recognition system.

7:30 **VI:25 COARSE-TO-FINE IMAGE SEARCH USING NEURAL NETWORKS**

CLAY D. SPENCE, JOHN C. PEARSON, and JIM BERGEN, David Sarnoff Research Center

The efficiency of image search can be greatly improved by using a coarse-to-fine search strategy with a multi-resolution image representation. However, if the resolution is so low that the objects have few distinguishing features, search becomes difficult. We show that search at such low resolutions can be made useful by using two techniques: 1) extract simple features at high-resolution and use them for searching at low-resolution, and 2) use context information, i.e., objects visible at low-resolution which are not the objects of interest but are associated with them. The use of multi-resolution search techniques also allows us to combine information about the appearance of the objects on many scales in an efficient way. We have illustrated these ideas by training a hierarchical system of neural networks to find clusters of buildings in aerial photographs of farmland.

THURSDAY AM
ORAL SESSION 9

ALGORITHMS & ARCHITECTURES

8:30 **O9.1 FINANCIAL APPLICATIONS OF LEARNING FROM HINTS (INVITED TALK)**
Y.S. ABU-MOSTAFA, California Institute of Technology

9:00 **O9.2 COMBINING ESTIMATORS USING NON-CONSTANT WEIGHTING FUNCTIONS**
VOLKER TRESP, Siemens AG, Central Research

In recent years there has been growing interest in the problem of combining estimators by weighted summation (mixing) where the weights are constant. We extend the mixing approach to include weighting functions which depend on the input. We show how these weighting functions can be derived from estimates of the performance of individual estimators. The approach is modular since the weighting functions can easily be modified (no retraining) if more estimators are added. The approach allows the incorporation of estimators which were not derived from data such as expert systems or algorithms.

9:20 **O9.3 AN INPUT OUTPUT HMM ARCHITECTURE**
YOSHUA BENGIO, Universite de Montreal and PAOLO FRASCONI, Universita di Firenze

We introduce a recurrent architecture having a modular structure and we formulate a training procedure based on the EM algorithm. The resulting model has similarities to hidden Markov models, but supports recurrent networks processing style and allows to exploit the supervised learning paradigm while using maximum likelihood estimation.

9:40 **O9.4 BOLTZMANN CHAINS AND HIDDEN MARKOV MODELS**
LAWRENCE SAUL and MICHAEL JORDAN, MIT

We develop a statistical mechanical framework for the modeling of discrete time series. In particular, we investigate the problem of maximum likelihood estimation in Boltzmann machines with m-state visible units, n-state hidden units, linear architectures, and periodic weights. We call these networks *Boltzmann chains* and show that they contain hidden Markov models (HMMs) as a special case. We also show how to implement the Boltzmann learning rule exactly, in polynomial time, without resort to simulated or mean-field annealing. The necessary computations are done by the transfer-matrix method, a general procedure for solving one-dimensional models in statistical mechanics. In general, Boltzmann chains can accomodate loops and bubbles and parameterize a larger family of probability distributions than HMMs.

10:00 BREAK

ORAL SESSION 10
ALGORITHMS & ARCHITECTURES

10:30 O10.1 BAYESIAN QUERY CONSTRUCTION FOR NEURAL NETWORK MODELS

GERHARD PAASS and JORG KINDERMANN, German National Research Center for Computer Science

If the collection of data is costly we can gain by actively selecting particular informative data points in a sequential way. In a Bayesian decision theoretic framework we develop a query selection criterion which explicitly takes into account the intended use of the model predictions. By Markov Chain Monte Carlo methods the necessary quantities can be approximated to desired precision. As the number of data points grows the model complexity is adapted by a Bayesian model selection strategy. The properties of a simplified version of the criterion are demonstrated in numerical experiments with MLP and RBF networks.

10:50 O10.2 USING A SALIENCY MAP FOR ACTIVE SPATIAL SELECTIVE ATTENTION:IMPLEMENTATION & INITIAL RESULTS

SHUMEET BALUJA and DEAN A. POMERLEAU, Carnegie Mellon University

In many vision based tasks, the ability to focus attention on the important portions of a scene is crucial for good performance on the tasks. In this paper we present a simple method of achieving spatial selective attention through the use of a saliency map. The saliency map indicates which regions of the input retina are important for performing the task. The saliency map is created through predictive auto-encoding. The performance of this method is demonstrated on a simple task which has multiple very strong distracting features in the input retina. Architectural extensions and application directions for this model are presented.

11:10 O10.3 MULTIDIMENSIONAL SCALING AND DATA CLUSTERING

THOMAS HOFMANN and JOACHIM BUHMANN, Rheinische Friedrich-Wilhelms-Universitat

Euclidian embedding and partitioning a data set which is characterized by pairwise dissimilarities of the data is a difficult combinatorial optimization problem. Algorithms for embedding such a data set in a Euclidian space, for clustering these data and for actively selecting data items to support the clustering process are discussed in the maximum entropy framework. The algorithms implement a new strategy for nonlinear dimension reduction and visualization. To yield a clustering solution of predefined quality, active data selection reduces the number of required data considerably.

11:30 O10.4 A NON-LINEAR INFORMATION MAXIMISATION ALGORITHM THAT PERFORMS BLIND SEPARATION

ANTHONY J. BELL and TERRENCE J. SEJNOWSKI, The Salk Institute

A new learning algorithm is derived which performs online stochastic gradient ascent in the mutual information between outputs and inputs of a network. In the absence of *a priori* knowledge about the 'signal' and 'noise' components of the input, propogation of information depends on calibrating network non-linearities to the detailed higher-order moments of the input pdfs. By minimising mutual information between outputs,as well as maximising their individual entropies, the network 'factorises' the input into Independent Components (ICA). As an example, we present near-perfect separation of five digitally mixed speech signals. Our simulations lead us to believe that our network performs better at blind separation than the 'H-J' network (Jutten & Herault, 1991), reflecting the fact that it is derived rigorously from the mutual information objective.

11:50 ADJOURN TO VAIL FOR WORKSHOPS

WORKSHOPS AT VAIL

DECEMBER 2, 1994

NOVEL CONTROL TECHNIQUES FROM BIOLOGICAL INSPIRATION

ORGANIZERS:Richard D. Braatz,(rdb@beethoven.che.caltech.edu), University of Illinois, James S. Schwaber,(schwaber@eplrx7.es.dupont.com), DuPont, David Touretzky,(dst@CS.CMU.EDU), Carnegie Mellon, Thomas F. Enders, Technical University Munich, K. P. Unnikrishnan,(unni@neuro.cs.gmr.com) General Motors

INTENDED AUDIENCE:Those interested in novel control techniques inspired from neurobiology.

Panel participants: Martin Pottmann, DuPont, Babatunde A. Ogunnaike, DuPont, James Keeler, MCC, Austin, Michael A. Henson, Louisiana State University, Gerald Dreyfus, ESPCI, Paris, Francis J. Doyle, Purdue

The well-known control theoretician Roger Brockett recently stated that profound advances in control theory may be achieved by developing a theory of control that sheds significant light on the neuroanatomy of at least one animal. The development of such a theory is clearly a very challenging problem and many important problems remain unsolved. The objective of this workshop is to overview some recent progress on developing novel control techniques inspired from a study of biological control systems, intermixed with ample time for discussion. Such issues for discussion may include but are not limited to: 1) is our current understanding of biological systems sufficient so that reverse-engineering their attributes of robustness, reliability, and nonlinear functional behavior is now practical? 2) just how novel are the control techniques described by the presenting authors? 3) what is the future potential?

A complementary workshop: "Open and Closed Problems in Neural Network Robotics" organized by Marcus Mitchell will be held on Saturday.

The presenters will represent three research groups with active research in the area of bio-control. This workshop promises to be thought-provoking with the aim of spending a substantial amount of thetime on discussions. A detailed schedule of the workshop follows

MORNING SESSION:

7:30 Dave Touretzky and A. David Redish summarize their cognitive neuroscience theory of rodent navigation with implications for hippocampal function, and its implementation on a mobile robot.

8:00 discussion period for Touretsky/Redish presentation

8:15 Thomas F. Enders and collaborators summarize their research efforts in using neural networks in the development of techniques for the scheduling, control,and on-

line optimization of batch fermentation processes (e.g. the alcoholic fermentation with yeast).

8:45 discussion period for Enders et al. presentation

9:00 panel/general discussion

AFTERNOON SESSION:

4:30 James S. Schwaber, Richard D. Braatz, Francis J. Doyle, Michael A. Henson, Martin Pottmann, and Babatunde A. Ogunnaike summarize their research efforts in developing novel process control techniques via inspiration from the cardiorespiratory reflexes.

5:00 discussion period for Schwaber et al. presentation

5:15 other workshop attendees present their work

6:00 panel/general discussion

MACHINE LEARNING APPROACHES IN COMPUTATIONAL MOLECULAR BIOLOGY

ORGANIZERS:Pierre Baldi (pfbaldi@juliet.caltech.edu),Soren Brunak (brunak@cbs.dth.dk)

INTENDED AUDIENCE:Researchers interested in the application of neural and other statistical methods to problems in Molecular biology.

A wealth of protein and DNA primary sequences is being generated by genome and other sequencing projects. Computational tools are increasingly needed to process this massive amount of data, to organise, compare and classify sequences, to detect weak patterns and similarities, to find and parse coding regions, to predict structure and function and reconstruct evolutionary trees. Sequence analysis problems have been tackled with classical statistical techniques, but also using artificial Neural Networks. Another trend in recent years, has been the casting of DNA and protein sequences problems in terms of formal languages using probabilistic automata, Hidden Markov Models and stochastic context free grammars. Machine learning techniques appear as a promising approach in this area.

This workshop will concentrate on the presentation and discussion of the most recent results on the application of machine learning approaches to problems in computational molecular biology. Emphasis will be both on methodological issues and biological relevance.

MORNING SESSION:

7:30 Pierre Baldi, "Hidden Markov Models of Human Genes"

8:00 Soren Brunak, "Construction of Low Similarity Data Sets of Sequences with Funtional Sites for Prediction Purposes"

8:30 Tim Hunkapiller

9:00 Anders Krogh, "Predicting Protein Secondary Structure with Structured Networks"

AFTERNOON SESSION:

4:30 Paul Stolortz, "Links between statistical physics and dynamic programming:applications to computational molecular biology"

5:00 Gary Stormo, "Neural Networks for the Identification of Functional Domains Common to Multiple Sequences"

5:30 Niels Tolstrup, "Neural Network Model of the Genetic Code"

6:00 Discussion

NOVELTY DETECTION AND ADAPTIVE SYSTEM MONITORING

ORGANIZERS:Thomas Petsche (petsche@scr.siemens.com) and Stephen J. Hanson (jose@learning.siemens.com), Siemens Corporate Research, Inc.; Mark Gluck (gluck@pavlov.rutgers.edu), Rutgers University

INTENDED AUDIENCE:Researchers interested in robot learning, exploration, and active learning systems in general.

Unexpected failure of a machine or system can have severe and expensive consequences. One of the most infamous examples is the sudden failure of military helicopter rotor gearboxes, which lead to acomplete loss of the helicopter and all aboard. There are many, more mundane, similar examples. The unexpected failure of a motor in a paper mill causes a loss of the product in production as well as lost production time while the motor is replaced. A computer or network overload, due to normal traffic or a virus invasion, can lead to a system crash that can cause loss of data and downtime.

In these examples and others, it can be cost effective to ``monitor" the system of interest and signal an operator when the monitored conditions indicate an imminent failure. This is analogous to periodically glancing at the fuel gauge in your car to make sure you do not run out of gas.

An adaptive system monitor, therefore, is an adaptive algorithm that estimates the condition of the system from a set of periodic measurements. This task is typically complicated by the fact that the measurements are complex and high dimensional. Adaptation is necessary since the measurements will depend on the peculiarities of the system being monitored and its environment.

This workshop will focus on the use of novelty detection for the problem of system monitoring. A novelty detector is a device or algorithm which is trained on a set of examples and learns to recognize or reproduce those examples. Any new example that is significantly different from the training set is identified as `novel" because it is unlike any example in the training set.

The purpose of the discussion is to bring together researchers working on different real world monitoring tasks and those working on novelty detection algorithms in order to hasten the development of broadly applicable adaptive monitoring algorithms.

We expect presentations on several applications areas involving a variety of novelty detection algorithms:

7:30 - 9:00 Helicopter gearbox monitoring presentations and discussions by RobertR. Kolesar (ONR), Kourosh Danai (U Mass), Peter Kazlas (U Colorado, Boulder) and Mark Gluck (Rutgers).

4:30 - 6:00 Engine and electric motor monitoring by Ken Marko (Ford), Scott Smith (Boeing), and Thomas Petsche (Siemens).

Recognizing novelty in classification tasks by Germano Vasconcelos University of Kent) and Dimitrios Bairaktaris (University of Stirling).

(This list of speakers is preliminary and subject to change.)

ANTHROPOMORPHIC SPEECH SIGNAL PROCESSING

ORGANIZERS: Hynek Hermansky (hynek@eeap.ogi.edu) and Misha Pavel (pavel@eeap.ogi.edu) Oregon Graduate Institue

INTENDED AUDIENCE: Practitioners in speech recognition, researchers interested inthe form and role of end-organ models.

Biologically faithful front-ends for speech and image tasks have seemed an attractive alternative to more traditional engineering representations - if we choose neural paradigms for recognition, then why not to look to biology for representation? With the availablity of silicon implementationsat reasonable cost, this enterprise would be expected to flourish.

Instead, more traditional representations often remain more effective. This workshop addresses why biologically faithful front-ends do not couple well to current neural-based recognizers. We will discuss biological front ends, and alternatives particularly representations based on traditional engineering practice but modified to include what is known about human perception. We will also consider in what circumstances biological front ends do offer an advantage, and explore what directions recognition technology must take to make better use of these models.

The Workshop will be oriented towards extensive discussions. Several potential participants have interests in presenting short talks to stimulate the discussions, among them:

MORNING SESSION:

7:30 Jont Allen (Bell Laboratories, Murray Hill), "Speech Recognition with Human Face"

8:00 Andreou Andreas (Johns Hopkins University), "Analog Auditory Models"

8:30 Malcom Slaney (Interval Research), "Correlograms"

9:00 Discussion

AFTERNOON SESSION:

4:30 Nelson Morgan (International Computers Science Institute and U C Berkeley),"Current Research in Stochastic Perceptual Auditory-event-based Models (SPAM) "

5:00 Chalapathy Neti (IBM Watson Center), "Neuromorphic speech processing for speech recognition in noisy environments."

COMPUTATIONAL ROLE OF LATERAL CONNECTIONS IN THE CORTEX

ORGANIZER:Joseph Sirosh , UT Austin

Intended Audience:Those interested in computation significance of connectivity patterns in cortex.

Substantial recent evidence indicates that intracortical connections develop in an activity-dependent manner much like the afferent connections to the cortex. For example, the pattern of long-range lateral connections is closely coupled to the pattern of feature detectors in the visual cortex, and can be altered by strabismus and visual deprivation. Several possible functions have been suggested for the lateral connections. They may (1) modulate receptive field properties in a context-dependent manner and mediate perceptual filling in, (2) mediate adult cortical plasticity such as dynamic receptive fields, (3) store associatory information such as Gestalt rules, (4) act as the substrate for stimulus-dependent synchronization and feature binding, and (5) form the locus of perceptual learning in the primary visual cortex.

The workshop will focus on collating the open questions and hypotheses about the functional role of intracortical connectivity, and formulating an agenda for computational and analytical modeling. How do patterned lateral connections form and develop? What do the patterns of lateral connectivity tell us about information stored in the cortex? How could associatory information in the lateral connections be expressed during cortical processing? How could lateral connections mediate learning processes in the cortex? What is their role in cortical plasticity? What types of neural network models are best suited for addressing such questions?

MORNING SESSION:

7:30 Gary Blasdel: Title to be announced.

8:00 Terrence Sejnowski: "Physiological Effects of Intrinsic Horizontal Connections in Visual Cortex"

8:30 Jack Cowan: "Geometric Visual Hallucinations and Lateral Cortical Connections"

AFTERNOON SESSION:

4:30 Shimon Edelman: "Computational models of 3D object representation in the visual cortex, and the possible role of lateral connections"

5:00 Jonathan Marshall: "Do lateral connections help stabilize perception during occlusion events?"

5:30 DeLiang Wang: "Lateral connections and coherent oscillations"

6:00 Joseph Sirosh: "Cooperative self-organization of lateral connections and feature detectors in the visual cortex"

UNSUPERVISED LEARNING RULES AND VISUAL PROCESSING

ORGANIZERS:Lei Xu (lxu@cs.cuhk.hk) and Laiwan Chan (lwchan@cs.cuhk.hk), The Chinese University of Hong Kong; Zhaoping Li (lwchan@cs.cuhk.hk), Hong Kong University of Science and Technology

There are three major types of unsupervised learning rules: competitive learning or vector quantization type, information preserving or *Principal Component Analysis (PCA)*} type, and the self-organizing topological map type. All of them are closely related to visual processing. For instance, they are used to interpret the developments of orientation and other feature selective cells, as well as development of cortical retinotopic maps such as ocular dominance and orientation columns.The development of the study of learning and the understanding of visual processing facilitate each other. Recent years, a number of advances have been made in both of the two areas.

For instance, in the area of unsupervised learning, (1) numerous algorithms for competitive learning, PCA learning, and self-organizing maps have been proposed; (2) several new theories and principles,like maximum coherence, minimum description length, finite mixtures with EM learning, statistical physics, Bayesian theory, exploratory projection pursuit, and local PCA, have been developed; (3) theories for unifying various unsupervised learning rules (e.g., multisets modeling learning theory) have been explored. In the area of visual processing, more knowledge is being gathered experimentally about how visual development can be preserved or altered by neural activities, neural transmitters/receptors, and the visual environment etc, providing the bases and constraints for various learning rules and motivating new learning rule studies. In addition, there has been more theoretical understandings on the dependence of the visual processing units on the visual input environment, supporting the rationality of unsupervised learning.

The purpose of this workshop is twofolds: (1) to summarize the advances on unsupervised learning and to discuss whether these advances can help the investigation on visual processing system; (2) to screening the current results on visual processing and to check if they can motivate or provide some hints on developing unsupervised learning theories. The targeted groups of participants are researchers working in either or both the area of learning and the study of visual processing.

MORNING SESSION 1:Chair, Lei Xu

7:30 John Wyatt and Ibrahim Elfadel (MIT), "Time-Domain Solutions of Oja's Equations"

7:50 Leon Bottou (Neuristique Paris) and Yoshua Bengio (University of Montreal), "Kmeans Performs Newton Optimization"

8:10 Lei Xu (The Chinese University of Hong Kong and Peking University), "Multisets Modeling Learning: An Unified Framework for Unsupervised Learning"

8:30 Nathan Intrator (Tel-Aviv University), "Information Theory Motivation For Projection Pursuit"

9:00 Peter Dayan (University of Toronto),"The Helmholtz Machine"

EVENING SESSION 1:Chair, Zhaoping Li

4:30 Juergen Schmidhuber (Technische Universitaet Muenchen), "Predictability Minimization And Visual Processing"

4:50 Tony Bell (Salk Institute), "Non-linear, Non-gaussian Information Maximisation: Why It's More Useful"

5:10 Zhaoping Li (Hong Kong University of Science and Technology),"Understanding The Visual Cortical Coding From Visual Input Statistics"

5:30 Klaus Obermayer (Universitaet Bielefel), "Formation Of Orientation And Ocular Dominance In Macaque Striate Cortex"

5:50 Joseph Sirosh (University of Texas at Austin), "Putative Functional Roles Of Self-organized Lateral Connectivity In The Primary Visual Cortex"

6:00 Discussion

MORNING SESSION 2:Chair, Laiwan Chan

7:30 Yoshua Bengio (University of Montreal),"Density Estimation with a Hybrid of Neural Networks and Gaussian Mixtures"

7:50 Eric Mjolsness (UCSD) and Steve Gold (Yale University), "Learning Object Models through Domain-Specific Distance Measures"

8:10 Dit-Yan Yeung (Hong Kong University of Science and Technology),"Auto-associative Learning of On-line Handwriting Using Recurrent Neural Networks"

8:30 Volker Tresp (Siemens AG, Central Research), "Training Mixtures of Gaussians with Deficient Data"

8:50 George F. Harpur and Richard W. Prager (Cambridge University), "A Fast Method for Activating Competitive Self-Organizing Neural-Networks"

EVENING SESSION 2:Chair, Lei Xu

4:30 Michael E. Hasselmo (Harvard University), "Neuromodulatory Mechanisms For Regulation Of Cortical Self-organization"

4:50 Sue Becker (McMaster University), "Learning To Cluster Visual Scenes With Contextual Modulation"

5:10 Jonathan A. Marshall (University of North Carolina at Chapel Hill),"Invisibility in Vision: Occlusion, Motion, Grouping, and Self-Organization"

5:30 Irwin King and Lei Xu (The Chinese University of Hong Kong), "A Comparative Study on Receptive Filters by PCA Learning and Gabor Functions"

5:50 Bernd Fritzke (Ruhr-Universitaet Bochum), "Detection of Visual Feature Locations with a Growing Neural Gas Network"

6:10 Discussion

STATISTICAL AND NEURAL NETWORK APPROACHES TO NATURAL LANGUAGE PROCESSING

ORGANIZERS:Gary Cottrell (gary@cs.ucsd.edu)

Recently there has been a great deal of activity in the Computational Linguistics community in applying statistical techniques to large text corpora. These techniques have been used for word sensedisambiguation, tagging of lexical items by their syntactic class, and for extracting frequent parse trees for faster parsing. At the same time, there has been a recognition among psycholinguists thatstatistical properties of sentences play an important role in the way that people process certain constructions.

Neural network models of natural language processing have mainlyfocused in recent years on lower-level processes, including learning of past tense constructions, pronunciation, and reading, although some approaches to parsing and learning of grammars have been attempted, with mixed results. In fact, the best results for larger grammars appear to have been achieved by hybrid approaches, while inductivelearning techniques have been most successful on small, restricted grammars.

FRIDAY MORNING:

Introductions

7:30 AM Mitch Marcus: "Statistical approaches to NLP"

8:00 AM Gary Cottrell: "Neural net approaches to NLP"Learning fsa's and pda's

Learning fsa's and pda's

8:30 AM Lee Giles "Learning a class of large finite state machines with a recurrent neural network"

8:50 AM Sreerupa Das "Differentiable symbol processing and an application to language induction"

Machine translation

9:10 AM Patrick Juola and James Martin: "Extraction of Transfer Functions through Psycholinguistic Principles"

FRIDAY AFTERNOON:

Parsing

4:30 PM George Berg "Single Network Approaches to Connectionist Parsing"

4:50 PM Ajay Jain, "PARSEC: Let Your Network do the Walking, but Tell it Where to Go."

5:10 PM Stan Kwasny: "Training SRNs to Learn Syntax"

5:30 PM Risto Miikkulainen "Parsing with modular networks"

Discussion

5:50 PM - 6:30 PM The assembled crew

SATURDAY MORNING

Word sense disambiguation/discovery/large text corpora

7:30 AM Hinrich Schuetze: "Unsupervised word sense disambiguation for improved text retrieval"

7:50 AM David Yarowsky "A comparison of word sense disambiguation algorithms"

8:10 AM Nick Chater "Neural networks as statistical inference: Why it's best to have all one's assumptions out in the open"

8:30 AM Eric Brill "Statistical language processing: What are numbers good for?"

Discussion

8:50 AM - 9:30 AM The assembled crew

SATURDAY AFTERNOON

Psycholinguistic modeling

4:30 PM Michael Gasser "Modular networks for language acquisition: Why and how"

4:50 PM David Plaut "Learning arbitrary and quasi-regular mappings in word reading with attractor networks"

5:10 PM Mark St. John "Practice makes perfect: The key role of construction frequency in sentence comprehension"

5:30 PM Kim Plunkett (unconfirmed), "Learning the Arabic plural: The case for minority default mappings in connectionist nets."

Discussion

5:50 PM - 6:30 PM The assembled crew

NEURAL NETWORKS IN MEDICINE

ORGANIZER:Paul E. Keller (pe_keller@gate.pnl.gov)

Intended Audience:People active or interested in applying neural networks in medicine.

Health care reform has become a major national focus. Among the many issues that have surfaced in the current health care debate, neural networks have the potential of being most beneficial in improving reliability and lowering cost.

The neural network approach in medical information processing offers many advantages including:

- rapid identification and diagnosis in real-time

- elimination of the impact of human fatigue and habituation on medical diagnosis

- automated or semi-automated analysis

- training by example.

The goal of this workshop is to investigate how neural networks can help improve the quality of health care and lower its cost. To accomplish this, the workshop will be a forum for researchers active in the field of medic al applications of neural networks to present their research and to participate in panel discussions. The panel discussions will be an opportunity for dialog among the workshop participants. Topics to be presented include pap smear analysis, cancer diagnosis, cancer screening, biomagnetic/ bioeletric signal processing, image segmentation, control of cardiac chaos, and predict ion and control of glucose metabolism. Topics of discussion will likely include clinical testing, reduction of false-negatives, how automation can lower health care costs, and the process of receiving government approval for medical products and procedures that incorporate neural network technology.

FRIDAY MORNING:

7:30 AM Optimizing networks for Atlas guided segmentation of brain images, Anand Rangarajan, Yale University

8:00 AM Neural Net Analysis of Solitary Pulmonary Nodules, Armando Manduca, Mayo Clinic

8:30 AM Using Neural Networks for Semi-automated Pap Smear Screening, Laurie Mango, MD, and James M. Herriman, Neuromedical Systems Inc.

9:00 AM Automated design of optical-morphological structuring elements for Pap smear screening, J. P. Sharpe, R. Narayanswamy, N. Sungar*, H. Duke, R. J. Stewart, L. McKeogh and K. M. Johnson, University of Colorado at Boulder and *California Polytechnic State University

FRIDAY AFTERNOON:

4:30 PM Comparing the prediction accuracy of statistical models and artificial neural networks in breast cancer, Harry Burke, MD, David Rosen, Phil Goodman, MD, New York Medical University and University of Nevada

5:00 PM Diagnosis of hepatoma by committee, Bambang Parmanto and Paul Munro, University of Pittsburg

5:30 PM Discussion

SATURDAY MORNING:

7:30 AM Neural Networks for Nonlinear Processing of Biomagnetic/Bioelectric Signals, Martin Schlang, Michael Haft, and Ralph Neuneier, Siemens

8:00 AM Neural networks distinguish demented subjects from elderly controls based on EEGs, Beatrice Golomb, MD, and Andrew F. Leuchter, MD, UCLA

8:30 AM Normal and Abnormal EEG Classification using Neural Networks and other techniques, Ah Chung Tsoi, University of Queensland

9:00 AM Issues in Controlling Cardiac Chaos, Gary W. Flake, Siemens Corporate Research

SATURDAY AFTERNOON:

4:30 PM Prediction and Control of the Glucose Metabolism of a Diabetic, Volker Tresp, John Moody* and Wolf-Ridiger Delong, Siemens and *Oregon Graduate Institute

5:00 PM Experiences in using neural networks for detecting coronary artery disease, Georg Doffner, Austrian Institute of Artificial Intelligence - University of Vienna

5:30 Panel Discussion

ADVANCES IN RECURRENT NETWORKS

ORGANIZERHava Siegelmann (iehava@ie.technion.ac.il):

Intended Audience:Those enamoured of, or frustrated with recurrent nets.

Unlike feedforward-acyclic networks, recurrent nets contain feedback loops, and thus give rise to dynamical systems. Theoretically, recurrent networks are very strong computationally. However, their dynamics introduces difficulties for learning and convergence.

This workshop will feature formal sessions, discussions, and a panel discussion aimed at understanding the dynamics, theoretical capabilities, and practical applicability of recurrent network. The panel discussion will focus on future directions of recurrent networks research.

FRIDAY MORNING:

Applications: Mahesan Niranjan (chair)

Lee A. Feldkamp (Remarks on Time-Lagged RNN--Training and Applications)

Jerry Connor (bootstrap methods in time series prediction)

Paul Muller (Programmable Aanlog Neural Computer: Design and Performance)

Lee Shung (Learning with smoothing Regularization)

Manuel Samuelides (application: design of neuro-filters)

Gary Kuhn (application of sensitivity analysis)

Morten With Pederson (Training and Pruning)

FRIDAY AFTERNOON:

Architectures: Lee Feldkamp (chair)

Paolo Frasconi - Learning and Rule Embedding

Lei Xu - Mixture Models and the EM Algorithm

Hava Siegelmann - Towards a Neural Language: Symbolic to Analog

General discussion

SATURDAY MORNING

Dynamics and Biology-Based Models: Pierre Baldi (chair)

Pierre Baldi - Trajectory Learning Using Shallow Hierarchies of Oscillators

Mahesan Niranjan - Stacking Multiple RNN Models of the Vocal Tract

Kenji Doya - Problems Concerning Bifurcations of Network Dynamics

Hugo deGaris - The CAM-Brain Project : Evolution of a Billion Neuron Brain

Dawei Dong - Associative Dynaimc Decorrelation

SATURDAY AFTERNOON

Fundamentals: Siegelmann (Chair)

Yoshua Bengio - On the Problem of Learning with Long-Term Dependencies

Barak Pearlmuter - On the Alleged Difficulty of Learning Long-Term Dependencies

Ricard Gavalda - On the Kolmogorov Complexity of RNN

Panel Discussion

DECEMBER 3, 1994

OPEN AND CLOSED PROBLEMS IN NEURAL NETWORK ROBOTICS

ORGANIZER:Marcus Mitchell (marcus@hope.caltech.edu)Chris M Bishop (Aston University)

Many of the presumed tenets of neural computation -- nonlinearity, parallelism, adaptation, real-time performance -- suggest that a "neuromorphic" approach to robotics problems could succeed where previous approaches have failed. Further, the amazing motor performance of humans and animals provides additional arguments for the potential benefits of "a sideways look" towards neurobiology. Spurred on by these and other factors, researchers from a variety of backgrounds have produced almost 15 years of research intended to elaborate a biologically-inspired robotics. This workshop will ask the questions "What has been accomplished so far?" and "What is to be done next?"

For all the research attempts to apply neural network ideas to robotics, it is still difficult to get clear answers to questions like "Can you use a neural network to control a 6 d.o.f. arm?" or "Do reinforcement learning and dynamic programming methods get killed by the curse of dimensionality?" In addition, robotics is an area with a vast and intimidating "non-neural" literature which must be considered. The main goal of this workshop is to stimulate discussion about what problems have been successfully attacked and what the most important current open problems entail. A secondary goal of the workshop is to produce a short consensus list of problem descriptions and their status.

A complementary workshop, titled "Novel Control Techniques from Biological Inspiration", organized by Jim Schwaber et al., may be of interest to participants. None of the presentations in that session will be on robotics, and its main focus will be on nonlinear dynamical systems, e.g. in chemical processes and in neural systems. It is a one day workshop to be held Friday.

MORNING SESSION:

7:30 - 7:35 Opening Remarks, Marcus Mitchell, Caltech

7:35 - 8:00 Why it's harder to control your robot than your arm: closed, open and irrelevant issues in inverse kinematics, Dave Demers, UCSD

8:05 - 8:30 Open Problem: Optimal Motor Hidden Units, Terry Sanger, JPL

8:35 - 9:00 Neural Network Vision for Outdoor Robot Navigation, Dean Pomerleau, CMU

AFTERNOON SESSION:

4:30 - 4:55 Learning New Representations and Strategies, Chris Atkeson, Georgia Tech

5:00 - 5:25 A Semi-Crisis for Neural Network Robotics:, Formal Specification of Robot Learning Tasks, Andrew Moore, CMU

5:30 - 6:30 Closing Discussion

NEURAL NETWORK ARCHITECTURES WITH TIME DELAY CONNECTIONS

ORGANIZERS:Andrew D. Back (back@elec.uq.oz.au), Eric A. Wan (ericwan@eeap.ogi.edu

INTENDED AUDIENCE: Researchers interested in the role of nonlinear feedforward structures that integrate elements of linear signal processing as an alternative to recurrent nets.

Nonlinear signal processing using neural network models is a topic of recent interest in various application areas. Recurrent networks offer a potentially rich and powerful modelling capability though may suffer from some problems in training. On the other hand, simpler network structures which have an overall feedforward structure, but draw more strongly on linear signal processing approaches have been proposed. The resulting structures can be viewed as a nonlinear generalizations of linear filters. This workshop is aimed at addressing issues surrounding networks which may be viewed in a nonlinear signal processing framework, focussing in particular on those which employ some form of time delay connections and generally limited recurrent connections. We intend to consolidate some of the recent theoretical and practical results, as well as addressing open issues.

MORNING SESSION:

7:30-7:45 Opening Discussion - Andrew Back, University of Queensland

7:45-8:00 "Computational Capabilities of Local-Feedback Recurrent Networks", Paolo Frasconi, University of Florence, Italy

8:00-8:15 " Issues in Representation: Recurrent Networks as Sequential Machines", C. Lee Giles and B.G. Horne, NEC Research Institute

8:15-8:30 "Properties of Recursive Memory Structures", Jose C. Principe, University of Florida

8:30-8:45 "A Local Model Net Approach to Modeling Nonlinear Dynamic Systems", Roderick Murray-Smith. MIT

8:45-9:15 Open forum: 5 minute presentations by participants

9:15-9:30 Question Time and Discussion

AFTERNOON SESSION:

4:30-4:45 "A Spatio-Temporal Approach to Visual Pattern Recognition", Lokendra Shastri, ICSI

4:45-5:00 "The Performance of Recurrent Networks for Classifying Time-Varying Patterns", Tina Burrows and Mahesan Niranjan, Cambridge University Engineering

Department

5:00-5:15 "Nonlinear Infomax With Adaptive Time Delays", Tony Bell, The Salk Institute

5:15-5:30 "The Sinc Tensor Product Network", Jerome Soller, University of Utah

5:30-5:45, "Discriminating Between Mental Tasks Using a Variety of EEG Representations", Chuck Anderson, Colorado State University

5:45-6:00 Open forum: 5 minute presentations by participants

6:00-6:30 Question Time and Closing Discussion

ALGORITHMS FOR HIGH DIMENSIONAL SPACES: WHAT WORKS AND WHY

ORGANIZER: MICHAEL P. PERRONE, (mpp@watson.ibm.com)

INTENDED AUDIENCE: The workshop is targeted on researchers interested in both theoretical and practical aspects of improving network performance..

The performance of certain regression algorithms is robust as the dimensionality of the data and parameter spaces are increased. Even in cases where the number of parameters is much larger than the number of data, performance is often robust. The central question of the workshop will be: What makes these techniques robust in high dimensions?.

High dimensional spaces have (asymptotic) properties that are nonintuitive when considered from the perspective of the two- and three-dimensional cases generally used for visual examples. Because of this fact, algorithm design in high dimensional spaces can not always be done by simple analogy with low dimensional problems. For example, a radial basis network is intuitively appealing for a one dimensional regression task; but it must be used with care for a 100 dimensional space and it may not work at all in 1000. Thus having a familiarity with the nonintuitive properties of high dimensional space may lead to the development of better algorithms.

We will discuss the issues that surround successful nonlinear regression estimation in high dimensional spaces and what we can do to incorporate these techniques into other algorithms and apply them in real-world tasks. The workshop will cover topics including the Curse of Dimensionality, Projection Pursuit, techniques for dimensionality reduction, feature extraction techniques, statistical properties of high dimensional spaces, local methods and all of the tricks that go along with these techniques to make them work.

MORNING SESSION:

7:30 "Statistical Properties of High Dimensional Spaces", Michael Perrone (IBM T.J. Watson Research Center)

8:00 "Computational Learning and Statistical Prediction", Jerome Friedman (Stanford

University)

8:30 "Discriminant Adaptive Nearest Neighbor Classification", Trevor Hastie and Rob Tibshirani (Stanford University)

9:00 "Local Methods in High Dimension: Are They Surprisingly Good But Miscalibrated?", David Rosen (New York Medical College)

AFTERNOON SESSION:

4:30 "Is There Anything Positive in High Dimensional Spaces?", Nathan Intrator (Tel Aviv University)

5:00 "Three Techniques for Dimension Reduction", John Moody (Oregon Graduate Institute)

5:30 "A Local Linear Algorithm for Fast Dimension Reduction", Nandakishore Kambhatla (Oregon Graduate Institute)

6:00 "Fuzzy Dimensionality Reduction", Yinghua Lin (Los Alamos National Lab)

DOING IT BACKWARDS: NEURAL NETWORKS AND THE SOLUTION OF INVERSE PROBLEMS

ORGANIZER: Chris M Bishop (Aston University)

INTENDED AUDIENCE: Researchers and practitioners in neural computing interested in inverse problems.

Many of the tasks for which neural networks are commonly used correspond to the solution of an `inverse' problem. Such tasks are characterized by the existence of a well-defined, deterministic `forward' problem which might, for instance, correspond to causality in a physical system. By contrast the inverse problem may be ill-posed, and may exhibit multiple solutions.

A wide range of different approaches have been developed to tackle inverse problems, and one of the main goals of the workshop is to contrast the way in which they address the underlying technical issues, and to identify key areas for future research. Ample time will be allowed for discussions.

MORNING SESSION:

7:30 "Welcome and overview" Chris Bishop (Aston)

7:35 "From ill-posed problems to all neural networks and beyond through regularization" Tomaso Poggio / Federico Girosi (MIT)

7:55 "Solving inverse problems using an EM approach to density estimation" Zoubin Ghahramani (MIT)

8:15 "Density estimation with periodic variables" Chris Bishop (Aston)

8:35 "Doing it forwards, undoing it backwards: high-dimensional compression and expansion" Russell Beale (University of Birmingham)

8:55 "Inversion of feed-forward networks by gradient descent" Alexander Linden (Berkeley)

9.15 Discussion

AFTERNOON SESSION:

4:30 "An iterative inverse of a talking machine" Sid Fels (Toronto)

4:50 "Diagnostic problem solving" Sungzoon Cho (Postech, S Korea)

5:10 "Multiple Models in Inverse Filtering of the Vocal Tract" M Niranjan (Cambridge)

5:30 "Goal directed model inversion" Silvano Colombano (NASA Ames)

5:50 "Predicting element concentrations in the SSME exhaust plume" Kevin Whitaker (University of Alabama)

6:10 Discussion

THE NEURAL BASIS OF LOCOMOTION: MODELS OF PATTERN GENERATORS

ORGANIZER: BARD ERMENTROUT

A
Abu-Mostafa, Y.S. 79
Ahmad, Subutai 63
Alexander, Jay A. 40
Alspector, J. 70
Ando, Hiroshi 53
Andreou, Andreas G. 70
Asogawa, Minoru 68

B
Back, Andrew D. 52
Baird, Leemon C. 31
Baldi, P. 67
Ballard, Dana H. 59
Baluja, Shumeet 80
Barkai, Edi 34
Barkai, N. 72
Barto, Andrew G. 27, 51
Becker, Suzanna 52
Behrman, Elizabeth C. 42
Bell, Anthony J. 81
Bengio, Yoshua 38, 39, 79
Bergen, Jim 78
Berke, Joshua 34
Berthold, Michael R. 37
Bischof, Horst 52
Bishop, Chris M. 55, 62
Bottou, Leon 39
Boyan, Justin A. 32
Bradtke, Steven J. 50
Brashers-Krug, Tom 30, 69
Brauer, Wilfried 44
Braver, Todd S. 74
Bregler, Christoph 78
Breiman, Leo 28
Brown, Thomas H. 33
Bruske, Jorg 56, 64
Buhmann, Joachim 80
Burgess, Neil 70
Burke, Harry B. 65
Burl, M. 67

C
Carnevale, Nicholas T. 33
Caruana, Rich 54, 62
Cauwenberghs, Gert 43
Chang, Eric I. 51
Chiang, Wan-Ping 59, 71
Choi, Young-Jae 71
Churcher, S. 42
Claiborne, Brenda J. 33
Coggins, Richard 55
Cohen, Jonathan D. 74
Cohn, David A. 63
Cooper, Leon N. 56
Cortes, Corinna 59, 71
Cottrell, Garrison W. 32, 39
Cowan, Jack D. 45
Crick, F.H. 29
Crites, Robert 51

D
Darrell, Trevor 60
Dayan, Peter 54, 68
Deco, Gustavo 44
Deffayet, Cedric 49
Diamond, Jay 37
Diorio, Chris 55, 71
Dong, Dawei W. 59, 78
Douglas, Rodney 29
Doya, Kenji 33, 48
Dreyfus, Gerard 68
Duff, Michael O. 50

E
Edelman, Shimon 47
Erwin, Ed 33, 48
Essa, Irfan 60

F
Fayyad, U.M. 67
Fels, S. Sidney 58
Feng, Jianfeng 73
Ferra, Herman 35
Fine, Terrence L. 73, 74
Finke, Michael 67
Flower, Barry 55
Frasconi, Paolo 38, 79
Fritzke, Bernd 61
Furlanello, Cesare 51

G
Ghahramani, Zoubin 41, 54, 61, 63, 69
Giles, C. Lee 37, 62, 63
Giuliani, Diego 51
Gold, Steven 53, 64
Goodman, Philip H. 65
Graham, Bruce 36

H
Hadjifaradji, Saeed 46
Hajto, J. 42
Hammer, Martin 33
Hammerstrom, Dan 26
Han, Il-Song 71
Hansen, Lars Kai 63
Harmon, Mance E. 31
Hasler, Paul 55, 71

Hasselmo, Michael E. 34
Hassibi, Babak 34
Hastie, Trevor 54
Haynes, Paul S. 55
Heil, Stefan 58, 66
Hinton, Geoffrey E. 54, 58, 61, 68, 77
Hochreiter, Sepp 37
Hofmann, Thomas 80
Holmes, A.J. 42
Horiuchi, Timothy 43
Horn, David 75
Horne, Bill G. 37, 62, 63
Hornik, Kurt 52
Huertas, J.L. 56

J
Jaakkola, Tommi 31, 32
Jabri, Marwan 37, 55
Jackel, L.D. 59, 71
James, Daniel L. 39
Jim, Kam 62
Jordan, Michael I. 31, 32, 41, 61, 63, 69, 79

K
Kailath, Thomas 34
Kairiss, Edward W. 33, 48
Kambhatla, Nanda 63
Kazlas, Peter T. 64
Ke, Liu 32, 42
Kearns, Michael 25
Kim, Ki-Chul 71
Kindermann, Jorg 80
Kitainik, L. 54
Klopf, A. Harry 31
Knerr, Stefan 68
Knierim, James J. 76
Koch, Christof 29
Kohonen, Teuvo 26
Kowalczyk, Adam 46
Krogh, Anders 35, 47
Kudrimoti, Hemant S. 76
Kuo, Jyh-Ming 73

L
Lee, Hwang-Soo 71
Lee, Yuchun 56, 66
Leen, Todd K. 35, 46, 63
Leerink, Laurens R. 37
Legleye, Claire 62
Lemmon, Michael 39
Linares-Barranco, B. 56
Lippe D. 70
Lippmann, Richard P. 51, 56, 66
Lu, Chien Ping 53

M
Maass, Wolfgang 34
Makhoul, John 58, 77
Malach, Raphael 47
Malaka, Rainer 33
Malpeli, Joseph G. 49
Manke, Stefan 67
Marchand, Mario 46
Marion, Glenn 44
Mathis, Donald W. 30
McCallum, R. Andrew 50
McNaughton, Bruce L. 76
McVey, Brian D. 32, 42
Mead, Carver 55, 71
Meir, Ronny 72
Miikkulainen, Risto 39, 47
Milgram, Maurice 54
Minch, Bradley A. 55, 71
Mjolsness, Eric 53, 64
Moore, Andrew W. 32
Movellan, Javier 58
Mozer, Michael C. 30, 40
Mukherjee, Sayandev 74
Murphy, Sean D. 33, 48
Murray, A.F. 42
Mussa-Ivaldi, Ferdinando 69

N
Nakano, Ryohei 38
Negishi, Michiro 30, 40
Neuneier, Ralph 63
Newsome, W.T. 33
Nix, David A. 56, 61
Niyogi, Partha 39
Nowlan, Steven J. 59

O
Obermayer, Klaus 33, 48
Ohira, Toru 45
Omohundro, Stephen M. 78
Oshima-Takane, Yuriko 40

P
Paass Gerhard 80
Pan, Hong 73
Pappu, Suguna 53
Pearson, John C. 78
Pedersen, Morten With 63
Pedroni, Volnei 43
Pentland, Alex 60
Perona, P. 67
Perrone, Michael P. 56
Personnaz, Leon 68
Pickard, Stephen 55

Pineda, Fernando J. 70
Platt, John C. 59
Pomerleau, Dean A. 58, 65, 80
Pouget, Alexandre 49, 75
Price, David 68
Principe, Jose C. 73

R
Ragg, Thomas 33
Raichle, Marc 28
Rangarajan, Anand 53, 64
Rao, Rajesh P.N. 59
Reggia, James A. 30, 41, 75
Revow, Michael 54, 68, 77
Robinson, A.J. 57
Rose, M.J. 42
Rosen, David B. 65
Roychowdhury, Vwani P. 73
Ruppin, Eytan 30, 41, 75

S
Saad, David 44, 72
Sackinger, Eduard 54
Sanger, Terence D. 32, 41
Saul, Lawrence 79
Schmidhuber, Jurgen 37, 58, 66
Schraudolph, Nicol N. 30, 38
Schulten, Klaus 48, 49
Schwartz, Anton 50
Schwartz, Richard 58, 77
Schwenk, Holger 54
Sejnowski, Terrence J. 30, 33, 38, 48, 49, 66, 75, 76, 81
Serrano, T. 56
Servan-Schreiber, David 74
Seung, H.S. 72
Shadmehr, Reza 30, 69
Shahian, Dr. David 56, 66
Shultz, Thomas R. 40
Simard, Patrice 54
Singh, Satinder P. 31, 32
Sirosh, Joseph 47
Skaggs, William E. 76
Skinner, Steven R. 42
Slaney, M. 57
Smith, Mike E.U. 55
Smyth, P. 67
Sollich, Peter 35, 44, 72
Sommer, Gerald 56, 64
Sompolinsky, H. 72
Spector, Kalanit Grill 47
Spence, Clay D. 78
Sperduti, Alessandro 62
Steck, James E. 42
Stensmo, Magnus 66
Stork, David G. 62
Suarez, Humbert 29
Sundareswaran, V. 59, 77
Sung, Kah Kay 39
Suzuki, Satoshi 53
Szymanski, Peter T. 39

T
Takane, Yoshio 40
Tenenbaum, Joshua B. 38
Terman, David 35
Thrun, Sebastian 36, 50, 66
Todd, Tom N. 55
Todorov, Emanuel 30
Todorov, Emanuel V. 38
Tokar, Robert L. 32, 42
Trentin, Edmondo 51
Tresp, Volker 63, 79
Trotman, David L. 55
Tsai, Kenneth Y. 33
Tsoi, Ah Chung 52
Tsung, Fu-Sheng 32, 39
Turmon, Michael J. 73
Tzonev, Svilen 49

U
Ueda, Naonori 38

V
Vaina, Lucia M. 59, 77
Vedelsby, Jesper 35, 47
Venkatesh, Santosh S. 45

W
Waibel, Alex 67
Wang, Changfeng 45
Wang, De Liang 35
Waterhouse, S.R. 57
Weigend, Andreas S. 56, 61, 64
Williams, Christopher K.I. 77
Willshaw David 36
Wolpert, Daniel M. 41, 69

X
Xu, Lei 61

Z
Zemel, Richard S. 76
Zhao, Ying 58, 77